

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2003, 00:43:55 ; Search time 6021 Seconds

(Without alignments)  
11334.675 Million cell updates/sec

Title: AF133659

Perfect score: 2345  
Sequence: 1 ATGGCGCTGCTCGCATGCA.....TATTAAAAAATCATCATTT 2345

Scoring table: IDENTITY NUC  
Gapop 10\_0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GeneBdb!:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_higo\_hum:\*  
40: em\_higo\_mus:\*  
41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID        | Description                |
|------------|--------|-------------|--------|--------------|----------------------------|
| 1          | 2345   | 100.0       | 2345   | 9 AF133659   | AF133659 Homo sapi         |
| 2          | 2340.2 | 99.8        | 2416   | 9 AB005289   | AB005289 Homo sapi         |
| 3          | 2332   | 99.4        | 2369   | 9 BC006323   | BC006323 Homo sapi         |
| 4          | 2328.2 | 99.3        | 2333   | 9 AF078777   | AF078777 Homo sapi         |
| 5          | 2248   | 95.9        | 2407   | 6 AF028561   | AF028561 Sequence          |
| 6          | 2248   | 95.9        | 2407   | 6 AR100662   | AR100662 Sequence          |
| 7          | 2150.4 | 91.7        | 2193   | 9 AK001418   | AK001418 Homo sapi         |
| 8          | 1833   | 78.2        | 2684   | 10 MMU43892  | MMU43892 Mus musculu       |
| 9          | 564.8  | 24.1        | 2979   | 3 AY051556   | AY051556 Drosophila        |
| 10         | 543.6  | 23.2        | 2187   | 8 AF287699   | AF287699 Arabidops         |
| 11         | 543.4  | 23.2        | 2373   | 8 AF360334   | AF360334 Arabidops         |
| 12         | 543.6  | 23.2        | 2037   | 8 AF287697   | AF287697 Arabidops         |
| 13         | 540.4  | 23.0        | 2452   | 8 AF287698   | AF287698 Arabidops         |
| 14         | 508.6  | 21.7        | 2043   | 8 AF287698   | AF287698 Arabidops         |
| 15         | 480.6  | 20.5        | 50701  | 2 AC018231   | AC018231 Drosophila        |
| 16         | 480.6  | 20.5        | 16682  | 3 AC093192   | AC093192 Drosophila        |
| 17         | 480.6  | 20.5        | 172175 | 3 AC010038   | AC010038 Drosophila        |
| 18         | 480.6  | 20.5        | 17583  | 3 AC105264   | AC105264 Drosophila        |
| 19         | 480.6  | 20.5        | 257867 | 3 AC005557   | AC005557 Drosophila        |
| 20         | 480.6  | 20.5        | 303191 | 3 AF003472   | AF003472 Drosophila        |
| 21         | 425.2  | 18.1        | 2262   | 8 SCATM1     | SCATM1 X82612 S. cerevisia |
| 22         | 425.2  | 18.1        | 2286   | 8 SCMDYGENE  | X81715 S. cerevisia        |
| 23         | 425.2  | 18.1        | 29286  | 8 SC9952X    | Z49212 S. cerevisia        |
| 24         | 418.8  | 17.9        | 282610 | 1 RPXX01     | AJ235270 Rickettsi         |
| 25         | 411.4  | 17.5        | 11621  | 1 AE013719   | AE013719 Yersinia          |
| 26         | 411.4  | 17.5        | 313050 | 1 AJ414152   | AJ414152 Yersinia          |
| 27         | 406    | 17.3        | 43227  | 2 TSF17      | AL049917 Arabidops         |
| 28         | 406    | 17.3        | 197655 | 8 ATCHRIV69  | AL161573 Arabidops         |
| 29         | 404.6  | 17.3        | 36310  | 8 SPAC8C9    | Z99168 S. pombe chr        |
| 30         | 361    | 15.4        | 4481   | 3 AF466308   | AF466308 Dictyoste         |
| 31         | 358.6  | 15.3        | 5369   | 6 A94856     | A94856 Sequence 3          |
| 32         | 347.8  | 14.8        | 43702  | 8 SPAC15A10  | Z97208 S. pombe chr        |
| 33         | 346.4  | 14.8        | 10086  | 1 AE009585   | AE009585 Brucella          |
| 34         | 335    | 14.3        | 9945   | 1 AE008036   | AE008036 Agrobacte         |
| 35         | 335    | 14.3        | 10322  | 1 AE009070   | AE009070 Agrobacte         |
| 36         | 310.4  | 13.2        | 346547 | 1 AP003012   | AP003012 Mesorhizo         |
| 37         | 308    | 13.1        | 2946   | 10 BC006634  | BC006634 Mus muscu         |
| 38         | 304.4  | 13.0        | 391    | 9 F241872S16 | AF241887 Homo sapi         |
| 39         | 304.4  | 13.0        | 100839 | 2 AC002417   | AC002417 Homo sapi         |
| 40         | 304.4  | 13.0        | 127243 | 9 AL359545   | AL359545 Human DNA         |
| 41         | 304.4  | 13.0        | 156799 | 2 AL359548   | AL359548 Homo sapi         |
| 42         | 304    | 13.0        | 2406   | 3 AF497513   | AF497513 Caenorhab         |
| 43         | 302.8  | 12.9        | 299350 | 1 SME591786  | AL591786 Sinorhizo         |
| 44         | 301.2  | 12.8        | 3008   | 3 AY084209   | AY084209 Drosophila        |
| 45         | 292    | 12.5        | 2848   | 10 RNABJ3004 | AJ003004 Rattus no         |

# ALIGNMENTS

RESULT 1  
AF133659  
LOCUS AF133659 2345 bp mRNA linear PRI 30-MAR-2001  
DEFINITION Homo sapiens ATP-binding cassette 7 iron transporter (ABC7) mRNA,  
complete cds; nuclear gene for mitochondrial product.

ACCESSION AF133659  
VERSION AF133659.1 GI:49277189

## KEYWORDS

SOURCE  
ORGANISM Homo sapiens.

REFERENCE  
1 (bases 1 to 2345)  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
Altkmetz, R., Raekind, W.H., Hutchinson, A., Schneck, N.D., Dean, M.  
and Koeller, D.M.

TITLE Mutation of a putative mitochondrial iron transporter gene (ABC7)  
in x-linked sideroblastic anemia and ataxia (XLSA/A)  
JOURNAL Hum. Mol. Genet. 8 (5), 743-749 (1999)  
MEDLINE 99214014  
PubMed 10196363  
REFERENCE 2 (bases 1 to 2345)  
AUTHORS Dean, M., Allikmets, R. and Hutchinson, A.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-1999) Human Genetics, National Cancer Institute,  
NCI-FCRDC, Frederick, MD 21703, USA

FEATURES  
Source Location/Qualifiers

1..2345  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xq13"  
1..2345  
/gene="ABC7"  
1..2259  
/gene="ABC7"  
/codon\_start=1  
/product="ATP-binding cassette 7 iron transporter"  
/protein\_id="AAD33045.1"  
/db\_xref="GI:4927190"  
/translation="MALAMHSMRWMAAAAFERRRSHAILRPLVSVSGSPQMRPH  
QLGALGTARAYQIESLSITWRLKGNSSQPLDAKALVWPLIEKRTCHHAGG  
GLHTDPKGLKDVDRKLIKAMLSYVWKDPPDRARVAISLGLGKAKNIVPFM  
PKVAVDSLNMGSMLNLSDA PNTVAATAVLIGYVSRAGA FPNENRVAIVGKVA  
ONSIRIRAKNVPFLHNIDLFHLSRQTGALSALDKDTRISTVLSLVVNLPIPH  
EVMISGVLVYKCAQAFALVTLGLGTATFVAVTRKRTRFRIEMNADNDANAI  
DSLNYETVKYFNNEERYEARYDGLKTYETASLSKSTLTMALNGSALISVGLTAI  
MLASOGIVAGTLIVGDLVMVNGLFLSLPLNPLGLVRETQALIDMNTLPIILKY  
DTQIKDKMASPLQTPOTATVAFNVMHFEYIEGQKVLISFEVPAGKVAIVGSG  
SGKSTIVLRFREYPOKSGSYLAGONIDVLSLEIRAVGVPODAVLFNITLYNL  
LYKNISAPREVVAVAKLAGADAILMPHXPDPVQGRGKISGGEKQVATIRALI  
KDPVYLIDEATSSLDSTTEBTITIGAMKDVAKHSTFIARLSLTVDADEITVDDG  
KVAERGTHGLANPHSIIYSSEMHPTQSSRVQNHNDPKWEAKENISKEERKLUDEI  
VNSVKGCGNCSC"

BASE COUNT 690 a 459 c 565 g 631 t  
ORIGIN

Query Match 100.0%; Score 2345; DB 9; Length 2345;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCTGCGGATGATCTTGCGGCGCGCGCGCGCGCTTTCGAAAAG 60  
DB 1 ATGGCGCTGCTGCGATGATCTTGCGCGCGCGCGCGCGCTTTCGAAAAG 60  
QY 61 CGCCGCACTCCGCAATTCGATCCGCGCTTCTGTTAGCGGCTCAGGTCGCGAG 120  
DB 61 CGCCGCACTCCGCAATTCGATCCGCGCTTCTGTTAGCGGCTCAGGTCGCGAG 120  
QY 121 TGGAGGCCACATCAACTCGGCGCTTGGGAAACCGCTCGAGCTTACCAATTCGAGGTCA 180  
DB 121 TGGAGGCCACATCAACTCGGCGCTTGGGAAACCGCTCGAGCTTACCAATTCGAGGTCA 180  
QY 121 TGGAGGCCACATCAACTCGGCGCTTGGGAAACCGCTCGAGCTTACCAATTCGAGGTCA 180  
DB 121 TGGAGGCCACATCAACTCGGCGCTTGGGAAACCGCTCGAGCTTACCAATTCGAGGTCA 180  
QY 181 TTTAAAGATACATGCGAGAGATTGGGAAAAGCAATTCAGGACAGTCTTCAATGCT 240  
DB 181 TTTAAAGATACATGCGAGAGATTGGGAAAAGCAATTCAGGACAGTCTTCAATGCT 240  
QY 241 GCAAAGGCTCTCCAGTATGCGCACTGATAGAAAAGAGACATGTTGGCATGCTATGCA 300  
DB 241 GCAAAGGCTCTCCAGTATGCGCACTGATAGAAAAGAGACATGTTGGCATGCTATGCA 300  
QY 301 GGAGGAGACTCTCACACAGACCCTAAAGAGGTTAAAGATGTTGATCTCGAAAAATC 360  
DB 301 GGAGGAGACTCTCACACAGACCCTAAAGAGGTTAAAGATGTTGATCTCGAAAAATC 360  
QY 361 ATTAAGCAATGCTTTCTTATGTGTGCCCCAAAGACAGGCGAGATCTACGAGCTAGAGTT 420  
DB 361 ATTAAGCAATGCTTTCTTATGTGTGCCCCAAAGACAGGCGAGATCTACGAGCTAGAGTT 420

QY 421 GCCATTTGCTGGGATTTTGGTGTGTGCAAAAGCCATGAATATGTGCTCCCTCATG 480  
DB 421 GCCATTTGCTGGGATTTTGGTGTGTGCAAAAGCCATGAATATGTGCTCCCTCATG 480  
QY 481 TTTAAATATGCTGTAGACAGGCTCAACAGATGTGCGGAAACATGCTGAACCTAGAT 540  
DB 481 TTTAAATATGCTGTAGACAGGCTCAACAGATGTGCGGAAACATGCTGAACCTAGAT 540  
QY 541 GCACCAATATACAGTTGCAACCATGCGCAACAGCACTTCTGATGGCTATGTATCAAGA 600  
DB 541 GCACCAATATACAGTTGCAACCATGCGCAACAGCACTTCTGATGGCTATGTATCAAGA 600  
QY 601 GCTGAGGCTGCTTTTAAAGAGTTGAAATGCAATGTTGCAAGTATGCCAGAT 660  
DB 601 GCTGAGGCTGCTTTTAAAGAGTTGAAATGCAATGTTGCAAGTATGCCAGAT 660  
QY 661 TCAATCCGAAAGATAGCCAAAATGTCTTCCATCTTCACAACTCGATCTGGGTTTT 720  
DB 661 TCAATCCGAAAGATAGCCAAAATGTCTTCCATCTTCACAACTCGATCTGGGTTTT 720  
QY 721 CACCTGAGCAGACAGACGGGAGCTTTATCTAAGGCTATTGACAGAGAACAAAGGGTATC 780  
DB 721 CACCTGAGCAGACAGACGGGAGCTTTATCTAAGGCTATTGACAGAGAACAAAGGGTATC 780  
QY 781 AGTTTGTCTGAGTGTCTTGTGTAATTAATCTTCCCATCATGTTTGAAGTATGCTT 840  
DB 781 AGTTTGTCTGAGTGTCTTGTGTAATTAATCTTCCCATCATGTTTGAAGTATGCTT 840  
QY 841 GTCACTGCTGTTTGTGTAATTAATGCGGTGCCAGTTTGTGTTGTTGTTGTTGTTGTT 900  
DB 841 GTCACTGCTGTTTGTGTAATTAATGCGGTGCCAGTTTGTGTTGTTGTTGTTGTTGTT 900  
QY 901 CTGGTATCATCACAGCATTCACAGTTGACAGTCAACAGTGTGGAATCTAGTTAGATA 960  
DB 901 CTGGTATCATCACAGCATTCACAGTTGACAGTCAACAGTGTGGAATCTAGTTAGATA 960  
QY 961 GAAATGAACAAAGAGATATGATGACAGTATGCTGCTATAGACTCACTGCTGAATAT 1020  
DB 961 GAAATGAACAAAGAGATATGATGACAGTATGCTGCTATAGACTCACTGCTGAATAT 1020  
QY 1021 GAACTGTGAAGTATTTAATATGAAGAATATGAACAGACAGATATGATGATTTTGG 1080  
DB 1021 GAACTGTGAAGTATTTAATATGAAGAATATGAACAGACAGATATGATGATTTTGG 1080  
QY 1081 AAGAGTATGAGACTGCTTCAATGAAGTCTCTACTGCTGCTATAGCTGAACTTTGGT 1140  
DB 1081 AAGAGTATGAGACTGCTTCAATGAAGTCTCTACTGCTGCTATAGCTGAACTTTGGT 1140  
QY 1141 CAAAGTCTATTTTCACTGCTGTTTAAACAGCTATTAATGCTGCTGCGCAGTCAGGAAAT 1200  
DB 1141 CAAAGTCTATTTTCACTGCTGTTTAAACAGCTATTAATGCTGCTGCGCAGTCAGGAAAT 1200  
QY 1201 GTGGCAGTATCCCTTATCTGTTGAGATCTAGTAATGTTGAAATGAACTTTTCACTT 1260  
DB 1201 GTGGCAGTATCCCTTATCTGTTGAGATCTAGTAATGTTGAAATGAACTTTTCACTT 1260  
QY 1261 TCATTAACCCCTGAATCTTCTGGAACCTGTATATGAGAGACTAGACAGACCTCATAGAT 1320  
DB 1261 TCATTAACCCCTGAATCTTCTGGAACCTGTATATGAGAGACTAGACAGACCTCATAGAT 1320  
QY 1321 ATGAACACCTTGTATCTACTCAAGGTAGACCCCAATTAAGAACAAGGTAGTGA 1380  
DB 1321 ATGAACACCTTGTATCTACTCAAGGTAGACCCCAATTAAGAACAAGGTAGTGA 1380  
QY 1381 TCTCCCTTCAATCAGACCAACAGACAGTACCGTGGCTTTGATATGTGCAATTTTGA 1440  
DB 1381 TCTCCCTTCAATCAGACCAACAGACAGTACCGTGGCTTTGATATGTGCAATTTTGA 1440  
QY 1441 TACATGAGGGCCAGAAAGTCTTAGTGAATATCTTTGAAGTCCCTGCGAGAAAGAA 1500  
DB 1441 TACATGAGGGCCAGAAAGTCTTAGTGAATATCTTTGAAGTCCCTGCGAGAAAGAA 1500  
QY 1501 GTGGCATTTGTAGAGGATGTGGGTACAGGGAAGAACAAATATGAGGCTATTATTTGCG 1560

```

Db 1501 GTGGCATTGAGAGGAGTGGGTCAAGGAAAGACAAATGAGGCTATTTATTCGC 1560
QY 1561 TTCTATGAGCCTCAAAAGGGTAGCAATTTATCTTGCTGCTCAAAATATATCAAGATGTGAGC 1620
Db 1561 TTCTATGAGCCTCAAAAGGGTAGCAATTTATCTTGCTGCTCAAAATATATCAAGATGTGAGC 1620
QY 1621 CTGAAAGCCTTCGAGAGGCGAGTGGAGTGTACCTCAGATGCTGCTCTCTCCATPAT 1680
Db 1621 CTGAAAGCCTTCGAGAGGCGAGTGGAGTGTACCTCAGATGCTGCTCTCTCCATPAT 1680
QY 1681 ACTATTTATTAACAACCTCTTATATGAAACATCACTGCTTACCTGAGAGAGTGTATGCA 1740
Db 1681 ACTATTTATTAACAACCTCTTATATGAAACATCACTGCTTACCTGAGAGAGTGTATGCA 1740
QY 1741 GTGGCAAAATTAGCTGAGCACTTCATGATGCAATCTTCGATGAGCACAATGATGACACC 1800
Db 1741 GTGGCAAAATTAGCTGAGCACTTCATGATGCAATCTTCGATGAGCACAATGATGACACC 1800
QY 1801 CAAGTAGGGGAAACAGAGACTCAAGCTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db 1801 CAAGTAGGGGAAACAGAGACTCAAGCTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1861 AGAGCCATTTTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 1861 AGAGCCATTTTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1921 TCGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1921 TCGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1981 ATTTTCATTTGACACAGATTTGTCAACAGTGTGTATGATGATGATGATGATGATGATGAT 2040
Db 1981 ATTTTCATTTGACACAGATTTGTCAACAGTGTGTATGATGATGATGATGATGATGATGAT 2040
QY 2041 CAGGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2041 CAGGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2101 TATTCAGAAATGTGGCATACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2101 TATTCAGAAATGTGGCATACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2161 GAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db 2161 GAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2221 GTCAATAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2221 GTCAATAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2281 TTTTGTGTTTGGATACATATTTGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Db 2281 TTTTGTGTTTGGATACATATTTGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
QY 2341 ACATT 2345
Db 2341 ACATT 2345

```

```

TITLE
Cloning and chromosomal mapping of a novel ABC transporter gene
(hABC7), a candidate for X-linked sideroblastic anemia with
spinocerebellar ataxia
JOURNAL
J. Hum. Genet. 43 (2), 115-122 (1998)
MEDLINE
98284536
REFERENCE
2 (bases 1 to 2416)
AUTHORS
Shimada, Y.
TITLE
Direct Submission
JOURNAL
Submitted (127-JUN-1997) Yoshikazu Shimada, Otsuka Pharmaceutical
Co. Ltd., Otsuka GSN Research Institute, Kagasuno, Kawachi-cho,
Tokushima, Tokushima 771-0192, Japan
(E-mail: shim@otsuka.genome.ad.jp, Tel: 81-886-65-2888,
Fax: 81-886-37-1035)
FEATURES
source
1..2416
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/locus_tag="hABC7"
/locus_type="placenta"
1..2416
/gene="hABC7"
69..2327
/codon_start=1
/product="ABC transporter 7 protein"
/protein_id="BAA28861.1"
/db_xref="GI:3228279"
/translation="MALLMHSWRMAAAAEKRRHSAILIRPLVSVSGSPQMPH
QUGALGTARAYIPESLSKITSYQRLKNGSGPFLDAKALQVPLEKRTCHGAGG
GHTDPKGLKQVDFRKIIKAMLSYWPEDRDLARVAISGFLGAGAAKVAIVPEM
FYAADVDSQMSNNMLNDAENTVAATVAATVAIIGVSRAGAAFEENAVNVEKVA
ONSIRIADNVEFLHNLDELGFLSRQALSKAIDRGKGISFVLSALVNLPIPF
EMLVSGVLYVACGAPALVLTGTATGTAFTVAATRMGTREIENKADNDAGNAAI
DSLAVETKYINRERYAORVDFPKLYETSLKSTSLAMNCGSAIPEVGLAI
MYLAQGIYAGTILYGDVYVNGLLPOLSLPNFTGYTREROLIDMTILFTLKV
DTQIDKRWASPLQITPQATVAFDNVHVEYLEGQVLSGISFEVPAKVAIVVEGS
SKSTSTVRLFRFEPPOKSIYLAQONIDVLSLRVAVVPODAVFNHTIYVNL
LYGNISASPEEYVAAKLAGLHDAILRMHGYDTQVGERGLKSGEKORVAIARIL
KDPVILVDHNTSSIDSTIEETILIGAMDVVGHRTSIFLAHLSLVYDADEITVDOG
KVAERGTGHGLLANPHSITSEMHITQSSRVQNDHPKWEAKENISKERKKLOBEI
VNSVKGCGKSC"
BASE COUNT
713 a 476 c 577 g 650 t
ORIGIN
Query Match 99.8%; Score 2340.2; DB 9; Length 2416;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2342; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGCGCTGCTCGCATGATCTTGGCGCTGGCGCGCGCGCTGTCGAAAAG 60
Db 69 ATGGCGCTGCTCGCATGATCTTGGCGCTGGCGCGCGCGCGCTGTCGAAAAG 128
QY 61 CGCGCGACTCGCGCATGATCTTGGCGCTTGTGCTGTATGCGGCTCGGAC 120
Db 129 CGCGCGACTCGCGCATGATCTTGGCGCTTGTGCTGTATGCGGCTCGGAC 188
QY 121 TGGAGGCAATCACTCGCGCGCTTGGAGCCGCTCGAGCCCTACAGATTCCAGAGTCA 180
Db 189 TGGAGGCAATCACTCGCGCGCTTGGAGCCGCTCGAGCCCTACAGATTCCAGAGTCA 248
QY 181 TTTAAAAGTATCATGAGAGAGATTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 249 TTTAAAAGTATCATGAGAGAGATTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
QY 241 GCAAGAGCTCTCCAGGTATGCGCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 309 GCAAGAGCTCTCCAGGTATGCGCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 368
QY 301 GGAAGAGAGCTCCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 369 GGAAGAGAGCTCCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428

```

|    |      |  |      |
|----|------|--|------|
| Db | 1509 | TACATTGAGGCCACGAAAGTCTTAGTGGAAATATCTTTGAAAGTCCCTGCAGGAAAGAAA     | 1568 |
| Qy | 1501 | GTGGCCATTGTAGAGAGGTAGTGGGTCAAGGAAAAGCACAATATGTAGGCTATTAATTTGCG   | 1560 |
| Db | 1569 | GTGGCCATTGTAGAGAGGTAGTGGGTCAAGGAAAAGCACAATATGTAGGAGCTATTAATTTGCG | 1628 |
| Qy | 1561 | TTCTATGTAGCCTCAAAAAGGTAGCATTTATCTTGGCTGCAAAATATACAAAGATGTAGC     | 1620 |
| Db | 1629 | TTCTATGTAGCCTCAAAAAGGTAGCATTTATCTTGGCTGCAAAATATACAAAGATGTAGC     | 1688 |
| Qy | 1621 | CTGAAAGCCTTTCGAGAGGCGAGTGGAGTGTACTCGATGAGTGTCTCTTCCATAT          | 1680 |
| Db | 1689 | CTGAAAGCCTTTCGAGAGGCGAGTGGAGTGTACTCGATGAGTGTCTCTTCCATAT          | 1748 |
| Qy | 1681 | ACTATTTATTAACAACCTCTTATATATGAAACATCAGTGTCTTCACTGAGAAAGTATGCA     | 1740 |
| Db | 1749 | ACTATTTATTAACAACCTCTTATATATGAAACATCAGTGTCTTCACTGAGAAAGTATGCA     | 1808 |
| Qy | 1741 | GTGGCAAAATTAAGTGTGACCTTCATGATGCAATTTCTTGGAAAGCCACATGATATGACACC   | 1800 |
| Db | 1809 | GTGGCAAAATTAAGTGTGACCTTCATGATGCAATTTCTTGGAAAGCCACATGATATGACACC   | 1868 |
| Qy | 1801 | CAAGTAGGGGGAACGAGACCTCAAGCTTTCAGAGAGAGAAAAGCAAAAGTAGTGAATTGCA    | 1860 |
| Db | 1869 | CAAGTAGGGGGAACGAGACCTCAAGCTTTCAGAGAGAGAAAAGCAAAAGTAGTGAATTGCA    | 1928 |
| Qy | 1861 | AGAGCACTTTTGAAGAGACCCGCCAGTCATCTATATGATGAAGTACTTCATCGTTAGAT      | 1920 |
| Db | 1929 | AGAGCACTTTTGAAGAGACCCGCCAGTCATCTATATGATGAAGTACTTCATCGTTAGAT      | 1988 |
| Qy | 1921 | TCGATTACTGAAGACATATCTTGTGTGCCATGAAAGATGTGTCTTCAACACAGAACTTCT     | 1980 |
| Db | 1989 | TCGATTACTGAAGACATATCTTGTGTGCCATGAAAGATGTGTCTTCAACACAGAACTTCT     | 2048 |
| Qy | 1981 | ATTTTCATTGACACACAGATGTGTCAACAGTGGTTGATGCATGAGAAATCATGTCTTGAT     | 2040 |
| Db | 2049 | ATTTTCATTGACACACAGATGTGTCAACAGTGGTTGATGCATGAGAAATCATGTCTTGAT     | 2108 |
| Qy | 2041 | CAGGTAAGGTAGGCCGACAGTGTGTAACCCACATGATGATTTGCTTGTAAACCTCATAGATC   | 2100 |
| Db | 2109 | CAGGTAAGGTAGGCCGACAGTGTGTAACCCACATGATGATTTGCTTGTGTAAACCTCATAGATC | 2168 |
| Qy | 2101 | TATTCAGAAATGTGGCATATCACAGACGACCGTGTGCAGAAACATGATTAACCCCAATG      | 2160 |
| Db | 2169 | TATTCAGAAATGTGGCATATCACAGACGACCGTGTGCAGAAACATGATTAACCCCAATG      | 2228 |
| Qy | 2161 | GAGCAAAAGAAAGAAATATATCCAAAGAGAGGAAAGAAAGAAACATCAAGAGAAAT         | 2220 |
| Db | 2229 | GAGCAAAAGAAAGAAATATATCCAAAGAGAGGAAAGAAAGAAACATCAAGAGAAAT         | 2288 |
| Qy | 2221 | GTCAATAGTGTGAAGGCTGTGAAACGTGTGTGTCTAAGTCAATTAAGACATTTTCTTT       | 2280 |
| Db | 2289 | GTCAATAGTGTGAAGGCTGTGAAACGTGTGTGTCTAAGTCAATTAAGACATTTTCTTT       | 2348 |
| Qy | 2281 | TTTTGTGTGTGTGACATATTTTGCCTGAGACAGAAATGTTTTATTAATAAAATAT          | 2340 |
| Db | 2349 | TTTTGTGTGTGTGACATATTTTGCCTGAGACAGAAATGTTTTATTAATAAAATAT          | 2408 |
| Qy | 2341 | ACATT 2345   |      |
| Db | 2409 | ACATT 2413   |      |

RESULT 3

LOCUS BC006323

DEFINITION Homo sapiens, ATP-binding cassette, sub-family B (MDR/TAP), member 7, clone MGC:112887 IMAGE:4138236, mRNA, complete cds.

ACCESSION BC006323

VERSION BC006323.1 GI:13623446

KEYWORDS MGC.

SOURCE Homo sapiens.



ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2369)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: ATCC  
DNA Library Preparation: Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: nisc\_mgc@hgrl.nih.gov  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantrop, S., Thomas, P.J., Tongsom, E.E., Touchman, J.W., Teurigon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAL Plate: 17 Row: K Column: 11  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.  
Location/Qualifiers  
1. 2369  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:12887 IMAGE:4138236"  
/tissue\_type="Muscle, Thadomysarcoma"  
/clone\_lib="NIH\_MGC\_17"  
/lab\_host="DH10B-R"  
/note="Vector: POT87"  
11. 2272  
/codon\_start=1  
/product="ATP-binding cassette, sub-family B (MDR/TAP), member 7"  
/protein\_id="AAH06323.1"  
/db\_xref="GI:13623447"  
/translation="MALIAMSWMAAAAAPEKRHSALIRPIVSVSGSGQWPRH QGALSTARAYQDIPESSKSTWRLKSGSGFDDAKALQWPLIBERTCHHGAG GSHATPKEGLKQVDTRKI I KAMLSYWPDRDLARVAISLGFGLKAKAMIVVF MFKYAVDSLQMGSMNLSDAPNTVAATAVAILGYVSRAGAFAFNEVRNAVFKV AQSIRIAANVFIHLNLDGFLSRQALSKALDRGRGISFVLSALFENLILIM FEYLVSVGLYKCGAOPALVTGTTGTATPAFAVATRRRTREIRIMNAGNAA IDSLAVETKYRNMRVNAQRVDYDFLKYETPASTLSTSLANLNGSARISVGLTA INVLASOGIVAGTLTVGDLVWVNGLLFQSLPLNLTGYRRETRQLIDMNTLFTLK VDTQIKDKWASPLQITPOTAVAFDNVHEVELEGKVLSGISFEVPAPKKAIVGGS GSGKSTVRLFRFEPKQSIYLAGQNTDVSLESLRAVAVPDDAVLFNHTIYNN LLYGNISASPEEYVAVALAGLHDAILRMPHGYDTQVGERGLSGSGEKORVAIAI LKDPVLIYDEATSSIDSTIEFTILGAMDVVRHSIFPAHRLSTVVDADETIVDO GKAEKCTHHGLLANPHSTISEMHTQSSRVQNDHPKBAKAKENISKEBERKLOE IYNSVKGCGSCC"  
BASE COUNT 701 a 465 c 572 g 631 t  
ORIGIN  
Query Match 99.4%; Score 2332; DB 9; Length 2369;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2345; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
QY 1 ATGGCGTGTCTCGGATGATCTTTCGCGCTGGCGCGCGCGCGCTGCTTCGAAAG 60

|||||  
Db 11 ATGGCGTGTCTCGGATGATCTTTCGCGCTGGCGCGCGCGCGCTGCTTCGAAAG 70  
QY CGCGGACATCGCGGATTCATCGCGGCTTTAGTCTCTGTAGCGGCTCAGGCGCAG 120  
Db CGCGGACATCGCGGATTCATCGCGGCTTTAGTCTCTGTAGCGGCTCAGGCGCAG 130  
QY 121 TGGAGGCAATCACTCGGCGCTTGGAGACCGCTCGAGCTAC--CAGATTCAGAG 177  
Db TGGAGGCAATCACTCGGCGCTTGGAGACCGCTCGAGCTACAGAGATTCAGAG 190  
QY 178 TCATTAAGATATCATATGAGAGATTTGGGAAAAGGCAATTCAGACAGTCTTAGAT 237  
Db TCATTAAGATATCATATGAGAGATTTGGGAAAAGGCAATTCAGACAGTCTTAGAT 250  
QY 238 GCGCAAGGCTCTCAGAGATGCGCAATGAGAAAGAGACATGTTGGATGATCAT 297  
Db GCGCAAGGCTCTCAGAGATGCGCAATGAGAAAGAGACATGTTGGATGATCAT 310  
QY 298 GCAGAGAGAGATCCACACAGACCCAAAGAGGTTAAAGATGTTACTCGGAAA 357  
Db GCAGAGAGAGATCCACACAGACCCAAAGAGGTTAAAGATGTTACTCGGAAA 370  
QY 358 ATCATTAAGCAATGCTTTCTTATGTGTGCGCCAAAGACAGCCCAATCTACAGCTAGA 417  
Db ATCATTAAGCAATGCTTTCTTATGTGTGCGCCAAAGACAGCCCAATCTACAGCTAGA 430  
QY 418 GTTGCATTTGCGGGAATTTTGGGTGGGCAAGGCCATGAATTTGGTTCCTTC 477  
Db GTTGCATTTGCGGGAATTTTGGGTGGGCAAGGCCATGAATTTGGTTCCTTC 490  
QY 478 ATGTTAAATATCTGTAGACAGCTCAACAGATTCGGGAAACATGCTGAACCTAGT 537  
Db ATGTTAAATATCTGTAGACAGCTCAACAGATTCGGGAAACATGCTGAACCTAGT 550  
QY 538 GATGCAACCAATACAGTTGCAACCATGGAACAGAGCTTGAATGGCTATGCTATCA 597  
Db GATGCAACCAATACAGTTGCAACCATGGAACAGAGCTTGAATGGCTATGCTATCA 610  
QY 598 AGAGCTGGAGCTCTTTTAAAGAGTTCGAAAGTGAATTTGGCAAGGTAGCCAG 657  
Db AGAGCTGGAGCTCTTTTAAAGAGTTCGAAAGTGAATTTGGCAAGGTAGCCAG 670  
QY 658 AATTCATCGAAGATAGCAAAATGCTTTCTGCATCTTCAACACTGATCTGGGT 717  
Db AATTCATCGAAGATAGCAAAATGCTTTCTGCATCTTCAACACTGATCTGGGT 730  
QY 718 TTTCACTGAGACAGACAGCGGAGCTTTATCTAAGCTATTGACAGAGAACAGGGGT 777  
Db TTTCACTGAGACAGACAGCGGAGCTTTATCTAAGCTATTGACAGAGAACAGGGGT 790  
QY 778 ATAGATTTTCTCTGAGTCTTGTGATTAATCTCTCCCATCATGTTGAAGGATG 837  
Db ATAGATTTTCTCTGAGTCTTGTGATTAATCTCTCCCATCATGTTGAAGGATG 850  
QY 838 CTGTGATGAGTCTTGTGATTAATCAATGCGGAGCCAGTTCCTTGGTAACCTTGA 897  
Db CTGTGATGAGTCTTGTGATTAATCAATGCGGAGCCAGTTCCTTGGTAACCTTGA 910  
QY 898 ACACTTGTATCATACAGAGCTTACAGATTCAGTCAACCGTGGAGAACTAGATTAGA 957  
Db ACACTTGTATCATACAGAGCTTACAGATTCAGTCAACCGTGGAGAACTAGATTAGA 970  
QY 958 ATGGAATGAACCAAGACATATGATGACAGATTAAGTCTGCTATAGACTACGCTGAAT 1017  
Db ATGGAATGAACCAAGACATATGATGACAGATTAAGTCTGCTATAGACTACGCTGAAT 1030  
QY 1018 TATGAACATGTAAGATTTTAAATGAAGATGAAGACAGAGATTTAGTGAATTT 1077  
Db TATGAACATGTAAGATTTTAAATGAAGATGAAGACAGAGATTTAGTGAATTT 1090  
QY 1078 TTGAAGCATGATGAGCTGCTTCAATGAAGATGACCTTACTGTGCTATGCTGAATTT 1137  
|||||

|    |      |  |      |
|----|------|--|------|
| Db | 1091 | TTGAAGACGTATGAGACTGCTTATTGAAAGTACTTACTCTGGCTATGCTGAACCTT       | 1150 |
| Qy | 1138 | GGTCAAAAGTGTATTTTCAGTGTGCGTTTAAACGTAATATGTTGCTGCCAGTCAAGGA     | 1197 |
| Db | 1151 | GGTCAAAAGTGTATTTTCAGTGTGCGTTTAAACGTAATATGTTGCTGCCAGTCAAGGA     | 1210 |
| Qy | 1198 | ATTGTGGCAGGTACCCCTTACTGTGTGGAGATCTATGTAATGTTGAATGACCTGTTTTTCAG | 1257 |
| Db | 1211 | ATTGTGGCAGGTACCCCTTACTGTGTGGAGATCTATGTAATGTTGAATGACCTGTTTTTCAG | 1270 |
| Qy | 1258 | CTTTCATTACCCCGTAACCTTTCGGGAACTGTATATAGAGACTATGACAAGCACTCATTA   | 1317 |
| Db | 1271 | CTTTCATTACCCCGTAACCTTTCGGGAACTGTATATAGAGACTATGACAAGCACTCATTA   | 1330 |
| Qy | 1318 | GATATGAACACTTGTATTACTTACTCAAGATGACACCCCAATTTAAAGACAAAGTGATG    | 1377 |
| Db | 1331 | GATATGAACACTTGTATTACTTACTCAAGATGACACCCCAATTTAAAGACAAAGTGATG    | 1390 |
| Qy | 1378 | GCATCTCCCTTCAGATCAACACACAGACGTAACCTGTGGCTTTGATTAATGTGATTTT     | 1437 |
| Db | 1391 | GCATCTCCCTTCAGATCAACACACAGACGTAACCTGTGGCTTTGATTAATGTGATTTT     | 1450 |
| Qy | 1438 | GAATATCTTGGAGGCGAGAAAGTCCCTTAGTGGAAATATCCCTTTGAAGTCCCTGCAGAAAG | 1497 |
| Db | 1451 | GAATATCTTGGAGGCGAGAAAGTCCCTTAGTGGAAATATCCCTTTGAAGTCCCTGCAGAAAG | 1510 |
| Qy | 1498 | AAATGGCCATTGTAGAGGTAGTGGGTCAAGGAAAGACAAATATGTAGGCTATTATTT      | 1557 |
| Db | 1511 | AAATGGCCATTGTAGAGGTAGTGGGTCAAGGAAAGACAAATATGTAGGCTATTATTT      | 1570 |
| Qy | 1558 | CGCTTCTATAGCCTCAAAAGGATAGCTTATCTTGTCTGTGCAAATATATCAAGATCTG     | 1617 |
| Db | 1571 | CGCTTCTATAGCCTCAAAAGGATAGCTTATCTTGTCTGTGCAAATATATCAAGATCTG     | 1630 |
| Qy | 1618 | AGCCTGGAAAAGCCTTCGGAGGGCAGTGGGAGTGGTACCTCAGGATAGCTGTCTCTTCAT   | 1677 |
| Db | 1631 | AGCCTGGAAAAGCCTTCGGAGGGCAGTGGGAGTGGTACCTCAGGATAGCTGTCTCTTCAT   | 1690 |
| Qy | 1678 | AATATCTATTTATTAACAACCTTATATATGAAACATCAGTGCCTTCACTGAGGAAGTGAT   | 1737 |
| Db | 1691 | AATATCTATTTATTAACAACCTTATATATGAAACATCAGTGCCTTCACTGAGGAAGTGAT   | 1750 |
| Qy | 1738 | GCACTGGCAAAATTAGCTGTGATTCATGATGCAATCTTTCGAATGCACTATGATATGAC    | 1797 |
| Db | 1751 | GCACTGGCAAAATTAGCTGTGATTCATGATGCAATCTTTCGAATGCACTATGATATGAC    | 1810 |
| Qy | 1798 | ACCCAAGTAGGGGGAACGAGGACCTCAAGCTTTCAGAGGAGAAAGCAAAAGATACCAAT    | 1857 |
| Db | 1811 | ACCCAAGTAGGGGGAACGAGGACCTCAAGCTTTCAGAGGAGAAAGCAAAAGATACCAAT    | 1870 |
| Qy | 1858 | GCAAGAGCAATTTTGAAGAGACCCGCCAGTATATCTCTATGATGAAGTACTTACCTGTTA   | 1917 |
| Db | 1871 | GCAAGAGCAATTTTGAAGAGACCCGCCAGTATATCTCTATGATGAAGTACTTACCTGTTA   | 1930 |
| Qy | 1918 | GATTGATTTACTGAAGAGACTATTTCTTGGTGCATGAGAGATGTGTCAACAGAACT       | 1977 |
| Db | 1931 | GATTGATTTACTGAAGAGACTATTTCTTGGTGCATGAGAGATGTGTCAACAGAACT       | 1990 |
| Qy | 1978 | TCATATTTTCATGACACAGATTTGTCAACAGTGTGTGATGACAGATGAATCATATGTCTTG  | 2037 |
| Db | 1991 | TCATATTTTCATGACACAGATTTGTCAACAGTGTGTGATGACAGATGAATCATATGTCTTG  | 2050 |
| Qy | 2038 | GATCAGGGTAGGTAGCGCGAAGCTGTATCCCAACATGTTTGTCTTACCCCTCATAGT      | 2097 |
| Db | 2051 | GATCAGGGTAGGTAGCGCGAAGCTGTATCCCAACATGTTTGTCTTACCCCTCATAGT      | 2110 |
| Qy | 2098 | ATCTAATTCAGAAATGTGGCATATACACAGAGCAGCGGTGTCAAGAACATATATACCCCAA  | 2157 |
| Db | 2111 | ATCTAATTCAGAAATGTGGCATATACACAGAGCAGCGGTGTCAAGAACATATATACCCCAA  | 2170 |
| Qy | 2158 | TGGGAGACAAAGAAAGAAATATATATCCAAAGAGAGAGAAAGAAAGAACTACAGAGAGAA   | 2217 |
| Db | 2171 | TGGGAGACAAAGAAAGAAATATATATCCAAAGAGAGAGAAAGAAAGAACTACAGAGAGAA   | 2230 |

|                       |   |  |                    |
|-----------------------|---|--|--------------------|
| Qy                    | 2218  | ATGTCGCAATAGTGGAAAGCGTGTGGAAACGTCGTGCGTCAATGACATTAAGACATTTTC | 2217               |
| Db                    | 2231  | ATTGTCATATGATGTGAAGGCGTGTGAAACGTCGTGCTAAGTCACATTAAGACATTTTC  | 2290               |
| Qy                    | 2278  | TTTTTTGGTTGTTTGGACTACATATTTTGACCTGAGCAGAAATGTTTTATTAAAAAAT   | 2337               |
| Db                    | 2291  | TTTTTTGTTGTTTGGACTACATATTTTGACCTGAGCAGAAATGTTTTATTAAAAAAT    | 2350               |
| Qy                    | 2338  | CATACATT 2345  |                    |
| Db                    | 2351  | CATACATT 2358  |                    |
| RESULT 4              |   |  |                    |
| LOCUS                 | AF078777  | 2333 bp  | mRNA               |
| DEFINITION            | Homo sapiens ABC transporter (ATM1) mRNA, partial cds; nuclear gene for mitochondrial product.  |  |                    |
| ACCESSION             | AF078777  |  |                    |
| VERSION               | AF078777.1  | GI:5702097   |                    |
| KEYWORDS              |   |  |                    |
| SOURCE                | Homo sapiens.   |  |                    |
| ORGANISM              | Homo sapiens  |  |                    |
| REFERENCE             | Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.   |  |                    |
| AUTHORS               | 1 (bases 1 to 2333)   |  |                    |
| TITLE                 | Ceere,P., Lill,R. and Kispal,G.   |  |                    |
| JOURNAL               | Identification of a human mitochondrial ABC transporter, the functional orthologue of yeast Atmdp   |  |                    |
| MEDLINE               | 99098366  |  |                    |
| PUBMED                | 9883897   |  |                    |
| REFERENCE             | 2 (bases 1 to 2333)   |  |                    |
| AUTHORS               | Ceere,P., Lill,R. and Kispal,G.   |  |                    |
| JOURNAL               | Direct Submission   |  |                    |
| TITLE                 | Submitted (16-JUL-1998) Cell Biology, University of Marburg, Robert Koch 5, Marburg 35033, Germany  |  |                    |
| FEATURES              | Location/Qualifiers   |  |                    |
| source                | 1..2333   |  |                    |
| gene                  | /organism="Homo sapiens"  |  |                    |
|                       | /db_xref="taxon:9606"   |  |                    |
|                       | <1..2333  |  |                    |
|                       | /gene="ATM1"  |  |                    |
|                       | <1..2247  |  |                    |
|                       | /gene="ATM1"  |  |                    |
|                       | /note="ATM1p"   |  |                    |
|                       | /product="ABC transporter"  |  |                    |
|                       | /codon_start=1  |  |                    |
|                       | /protein_id="A047141.1"   |  |                    |
|                       | /db_xref="GI:5702098"   |  |                    |
|                       | /translation="AMHSMWMAAAAFEFKRHSAILIRPLVSSGSGPOMRPHOLGALGRATVQIPESLKITWRGLKNGSGQFLDAKALQWPLIEKRTKTHHAGGLHTDPEGLDQVDVTRKIRKMSVYMKRPPDRIKALVSLGLGKAKNIVYPPFKTAVDSIDNSGNNMLSDAPNTVAATMAVALIGVSRGAAFPENVRNAVGAQNSIRRAKQVFLHNLMDIGFLSRQTALSKAIDGTRGISEPLSALVNLPIPEFWMLVSGVLKYKCAQFALVTLGLGTGTAVTAVTRMTRFRIEMNKADNDAGNAALDISLLNRYEVKFNENRVEAORVDEGLFTYERASLKSSTGLAMFGOSAIPSVGLTAIMVLAASQIVAGTLVGDVWVWNGLLFOLSLPMLNLTGRTSGLRSTALIDMTLFTLLKVDQIHKDMVASPLQITPOTATVADNVHFERITBEQKPLSGSEVPACKVAIVGSGSGSSTVRLFLFRFEPQKSGTYLAGQNIQDSVLSLRAGVAVPQDVLPHNTTYNLVGNISAPSEPVAVAKLAGLHDAILRMPHGYDQVBERGLKSGEKQRAVALAIRLIKQPPVILVDEATSLSDITTEETLIGAMKDVVKHRTSIFIAHRLSTVDADETIVLDGKVAERGNHGLANPHSIVSEPMHTQSSRVQNHNDPNMWEAKENISKEERKKLOEIVNSVKGCGNGSC" |  |                    |
| BASE COUNT            | 688 a   | 456 c  | 560 g 629 t        |
| ORIGIN                |   |  |                    |
| Query Match           | 99.3%;  | Score 2328.2;  | DB 9; Length 2333; |
| Best Local Similarity | 99.9%;  | Pred. No. 0;   |                    |
| Matches 2330;         | Conservative 0;   | Mismatches 3;  | Indels 0; Gaps 0;  |
| Qy                    | 13  | GCAGTCATCTTGGCGCTGGGCGCGCGCGGCGCTGCTTTGCAAAAGCCGCGCATCTCC    | 72                 |

|||||  
Db 1 GCGATGCAATCTTGGCCGTGGGCGGCGGCGGCTGCTTTCGAAAAGCCGCGCACCTCC 60  
Qy 73 GCGATTCGATCCGGCCCTTTAGTCTCTGTGTAGCGGCTCAGGTCGCGAGTGGAGGCCACAT 132  
Db 61 GCGATTCGATCCGGCCCTTTAGTCTCTGTGTAGCGGCTCAGGTCGCGAGTGGAGGCCACAT 120  
Qy 133 CAATCGGCGGCTTTGGGAAACCGCTCGAGCTTACGAGTTCAGAGTCAATTAAGATATC 192  
Db 121 CAATCGGCGGCTTTGGGAAACCGCTCGAGCTTACGAGTTCAGAGTCAATTAAGATATC 180  
Qy 193 ACATGGCAGAGATTGGGAAAAAGGCAATTCAGAGCAAGTCTTAAGTGTGTGCAAGGCTCTC 252  
Db 181 ACATGGCAGAGATTGGGAAAAAGGCAATTCAGAGCAAGTCTTAAGTGTGTGCAAGGCTCTC 240  
Qy 253 CAGGTATGGCCATGATAGAAAAGGGAATGTTGGCATGGTCAATGACGAGGAGGAGCTC 312  
Db 241 CAGGTATGGCCATGATAGAAAAGGGAATGTTGGCATGGTCAATGACGAGGAGGAGCTC 300  
Qy 313 CACACAGACCCCAAAAGAGGTTAAAGATGTGATATCTCGAAAATCATATAAGCAATG 372  
Db 301 CACACAGACCCCAAAAGAGGTTAAAGATGTGATATCTCGAAAATCATATAAGCAATG 360  
Qy 373 CTTTCTTATGTGTGGCCCAAAAGACAGGCGCAGATCTACGAGCTTAAGATTGCCCTG 432  
Db 361 CTTTCTTATGTGTGGCCCAAAAGACAGGCGCAGATCTACGAGCTTAAGATTGCCCTG 420  
Qy 433 GGATTTTGGGTGGTGCAAAAAGGCGCATGAAATATGTTGGTCCCTTCATGTTTAAATATGCT 492  
Db 421 GGATTTTGGGTGGTGCAAAAAGGCGCATGAAATATGTTGGTCCCTTCATGTTTAAATATGCT 480  
Qy 493 GTAGACAGCCTCAACAGATGTTCGGGAAACATGCTGAACCTGAATGATGACCAAAATACA 552  
Db 481 GTAGACAGCCTCAACAGATGTTCGGGAAACATGCTGAACCTGAATGATGACCAAAATACA 540  
Qy 553 GTTGCAACCATGGCAACAGCAGTCTGTGATGGCTATGGTGTATCAAGCTGGAGCTGCT 612  
Db 541 GTTGCAACCATGGCAACAGCAGTCTGTGATGGCTATGGTGTATCAAGCTGGAGCTGCT 600  
Qy 613 TTTTAAACGAAAGTTCAAAAATGCAATTTGGCAAGGATGCGCCGAATTCATCCGAAAGA 672  
Db 601 TTTTAAACGAAAGTTCAAAAATGCAATTTGGCAAGGATGCGCCGAATTCATCCGAAAGA 660  
Qy 673 ATAGCCAAAAATGTCTTCTTCATCTTCAACAACCTGATCTGGGTTTTTCACTGAGAGA 732  
Db 661 ATAGCCAAAAATGTCTTCTTCATCTTCAACAACCTGATCTGGGTTTTTCACTGAGAGA 720  
Qy 733 CAGACGGGAGCTTTATCTAAGGCTATTGACAGAGAAACAGGGGTATCAGTTTTGTCTG 792  
Db 721 CAGACGGGAGCTTTATCTAAGGCTATTGACAGAGAAACAGGGGTATCAGTTTTGTCTG 780  
Qy 793 AGTGCCTTGGATTTTAAATCTTCTTCCCATCATGTTTGAAGGATGCTTGTCAAGTGTGTT 852  
Db 781 AGTGCCTTGGATTTTAAATCTTCTTCCCATCATGTTTGAAGGATGCTTGTCAAGTGTGTT 840  
Qy 853 TTGTATTACAAAATCGGGTGCACAGTTGCTTTGGTAAACCTTGGAAACCTTGATACATAC 912  
Db 841 TTGTATTACAAAATCGGGTGCACAGTTGCTTTGGTAAACCTTGGAAACCTTGATACATAC 900  
Qy 913 ACAGCATTCACAGTTGACGTCAACGCGTGGAGAACTAGATTAGAAATAGAAATGAAACAAA 972  
Db 901 ACAGCATTCACAGTTGACGTCAACGCGTGGAGAACTAGATTAGAAATAGAAATGAAACAAA 960  
Qy 973 GCGATATATATGAGATATGCTGCTATATAGACTCTGCTGAATTAATGAAACGTGTAAAG 1032  
Db 961 GCGATATATATGAGATATGCTGCTATATAGACTCTGCTGAATTAATGAAACGTGTAAAG 1020  
Qy 1033 TATTTATATATGAAAGATATGAGACACAGATATGATGGAATTTTGAAGACGTATGAG 1092  
Db 1021 TATTTATATATGAAAGATATGAGACACAGATATGATGGAATTTTGAAGACGTATGAG 1080  
Qy 1093 ACTGCTTCATTGAAAGTACCTTACTCTGGCTATGCTGAACCTTGGTCAAAAGTGTATT 1152  
|||||

Db 1081 ACTGCTTCATTGAAAGTACCTTACTCTGGCTATGCTGAACCTTGGTCAAAAGTGTATT 1140  
Qy 1153 TTGAGTGTGGTTTAAACAGCTATATATGTTGCTCGCCAGTCAGGAAATTTGGCAGTACC 1212  
Db 1141 TTGAGTGTGGTTTAAACAGCTATATATGTTGCTCGCCAGTCAGGAAATTTGGCAGTACC 1200  
Qy 1213 CTTACTGTGGAATCTATGTAATGTGTAATGACCTGCTTTTTCAGCTTTCATTAACCCCTG 1272  
Db 1201 CTTACTGTGGAATCTATGTAATGTGTAATGACCTGCTTTTTCAGCTTTCATTAACCCCTG 1260  
Qy 1273 AACTTTTGGGAATCTGTAATATAGAGAACTAGCAAGCACTCATAGTAATGAACACTTG 1332  
Db 1261 AACTTTTGGGAATCTGTAATATAGAGAACTAGCAAGCACTCATAGTAATGAACACTTG 1320  
Qy 1333 TTTACTCTACTCAAGGTAGACACCCAAATTAAGAACAAAGTGAATGGCATCTCCCTTCAG 1392  
Db 1321 TTTACTCTACTCAAGGTAGACACCCAAATTAAGAACAAAGTGAATGGCATCTCCCTTCAG 1380  
Qy 1393 ATCACACACAGACAGCTACCGTGCCCTTGTATATGTGCAATTTTGAATCATTTGAGGGC 1452  
Db 1381 ATCACACACAGACAGCTACCGTGCCCTTGTATATGTGCAATTTTGAATCATTTGAGGGC 1440  
Qy 1453 CAGAAAGTCTTATGTGGAATATCTTTGAAGTCCCTGACAGAAAGAAAGTGGCCATTGTA 1512  
Db 1441 CAGAAAGTCTTATGTGGAATATCTTTGAAGTCCCTGACAGAAAGAAAGTGGCCATTGTA 1500  
Qy 1513 GGAAGTATGAGGTCAAGGAAAGACAAATATGAGAGCTATATTTGGCTCTATGAGCCT 1572  
Db 1501 GGAAGTATGAGGTCAAGGAAAGACAAATATGAGAGCTATATTTGGCTCTATGAGCCT 1560  
Qy 1573 CAAAAGGTAGACATTAATCTTGTGTCAAAATATACAAAGTGTAGCCCTGAAAGCCTT 1632  
Db 1561 CAAAAGGTAGACATTAATCTTGTGTCAAAATATACAAAGTGTAGCCCTGTAAGCCTT 1620  
Qy 1633 CGAAGGCAATGGGAGTGGATCTCTCAGAGTGTGTCTCTTCCATATATATTTATTTAC 1692  
Db 1621 CGAAGGCAATGGGAGTGGATCTCTCAGAGTGTGTCTCTTCCATATATATTTATTTAC 1680  
Qy 1693 AACCTCTTATATGGAACAAATCAGTGGCTTCACTGAGGAAGTGTATGAGTGGCAAAATTA 1752  
Db 1681 AACCTCTTATATGGAACAAATCAGTGGCTTCACTGAGGAAGTGTATGAGTGGCAAAATTA 1740  
Qy 1753 GCTGCACTTCATGATGCAATTTCTTGCATATGCCATGATATATGACCCAAATGAGGAA 1812  
Db 1741 GCTGCACTTCATGATGCAATTTCTTGCATATGCCATGATATATGACCCAAATGAGGAA 1800  
Qy 1813 CGAGCACTCAAGCTTTCAGAGGAGAAAGCAAAAGATGACAAATTCAGAGCCATTG 1872  
Db 1801 CGAGCACTCAAGCTTTCAGAGGAGAAAGCAAAAGATGACAAATTCAGAGCCATTG 1860  
Qy 1873 AAGGACCCCCCAATCTATCTTATGTAAGGCTACTTCAATCCGTATATGATTAATCTGAA 1932  
Db 1861 AAGGACCCCCCAATCTATCTTATGTAAGGCTACTTCAATCCGTATATGATTAATCTGAA 1920  
Qy 1933 GAGACTATCTTGGTGCATGAAGGATGTGTCAAAACACAGAACTTCTATTTCAATTGCA 1992  
Db 1921 GAGACTATCTTGGTGCATGAAGGATGTGTCAAAACACAGAACTTCTATTTCAATTGCA 1980  
Qy 1993 CACAGATTGTCAACAGTGTGTATGCAATGAATCAATGTCTTGAATCAGGTTAAGTGA 2052  
Db 1981 CACAGATTGTCAACAGTGTGTATGCAATGAATCAATGTCTTGAATCAGGTTAAGTGA 2040  
Qy 2053 GCGGAACGTGTGATCCCAACAATGGTTTGTGTGTAACCTCATAGTATCTATTCAGAAATG 2112  
Db 2041 GCGGAACGTGTGATCCCAACAATGGTTTGTGTGTAACCTCATAGTATCTATTCAGAAATG 2100  
Qy 2113 TGGCATATCAGAGACCGGTGTGAGAAACATGATTAACCCCAATGGGAAAGCAAGAAA 2172  
Db 2101 TGGCATATCAGAGACCGGTGTGAGAAACATGATTAACCCCAATGGGAAAGCAAGAAA 2160  
Qy 2173 GAAATATATCCAAAGAGAGGAAAGAAAGAACTACAAAGAAATTTGCAATAGTGTG 2232  
Db 2161 GAAATATATCCAAAGAGAGGAAAGAAAGAAAGAACTACAAAGAAATTTGCAATAGTGTG 2220  
|||||

2233 AAAGCTGTGAAAGCTTGTGCTAAGTACATAGACATTTCTTTTGTGTTT 2292  
Db 2221 AAAGCTGTGAAAGCTTGTGCTAAGTACATAGACATTTCTTTTGTGTTT 2280  
Qy 2293 GGACTACATATTTGACATGAAAGCAGATTTGTTTATTAATAAATCATACATT 2345  
Db 2281 GGACTACATATTTGACATGAAAGCAGATTTGTTTATTAATAAATCATACATT 2333

RESULT 5  
AR028561 2407 bp DNA linear PAT 29-SEP-1999  
LOCUS AR028561  
DEFINITION Sequence 2 from patent US 5858719.  
ACCESSION AR028561  
VERSION AR028561.1 GI:5940534  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2407)  
AUTHORS Hillman,J.L., Shah,P. and Corley,N.C.  
TITLE Polynucleotides encoding human Atp binding-cassette transport  
JOURNAL Patent: US 5858719-A 2 12-JAN-1999;  
FEATURES  
source location/Qualifiers  
1..2407  
/organism="unknown"

BASE COUNT 726 a 477 c 570 g 629 t 5 others  
ORIGIN

Query Match 95.9%; Score 2248; DB 6; Length 2407;  
Best local similarity 99.4%; Pred. No. 0; Mismatches 13; Indels 1; Gaps 1;  
Matches 2265; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

Qy 8 TGTCTGCGATGATCTTGGCGCTGGCGCGCGCGCGCTTTCGAAAGCCCGGC 67  
Db 1 TGTCTGCGATGATCTTGGCGCTGGCGCGCGCGCTTTCGAAAGCCCGGC 60  
Qy 68 ACTCCGCGATTTGATCCGCGCTTTAGTCTCTGTTAGCGGCTCAGTCCGCGTGAAGC 127  
Db 61 ACTCCGCGATTTGATCCGCGCTTTAGTCTCTGTTAGCGGCTCAGTCCGCGTGAAGC 120  
Qy 128 CACATCAACTCGCGCGCTTGGGAAACGCTCGAGCTACACAGATTCGAGATTAATAA 187  
Db 121 CACATCAACTCGCGCGCTTGGGAAACGCTCGAGCTACACAGATTCGAGATTAATAA 180  
Qy 188 GTATCAGATGCGAGATTTGGGAAAGGCAATTCAGACAGTTCCTAGATGCTGCAAG 247  
Db 181 GTATCAGATGCGAGATTTGGGAAAGGCAATTCAGACAGTTCCTAGATGCTGCAAG 240  
Qy 248 CTCTCCAGGTATGCGCATGATGAAAGAGACATGTTGGCATGCTCATGAGAGAG 307  
Db 241 CTCTCCAGGTATGCGCATGATGAAAGAGACATGTTGGCATGCTCATGAGAGAG 300  
Qy 308 GACTCCACACAGACCCAAAGAGGTTAAAGATGTTGATCTCGGAAATCATATAAG 367  
Db 301 GACTCCACACAGACCCAAAGAGGTTAAAGATGTTGATCTCGGAAATCATATAAG 360  
Qy 368 CAATCTTTCTATGTGTGGCCCAAGACAGGCTCAGATTCAGACTAGAGTTGCATT 427  
Db 361 CAATCTTTCTATGTGTGGCCCAAGACAGGCTCAGATTCAGACTAGAGTTGCATT 420  
Qy 428 CGCTGGGATTTTGGGTGTGCAAAAGCCATGATATTTGGTCTCCCTCATGTTTAAT 487  
Db 421 CGCTGGGATTTTGGGTGTGCAAAAGCCATGATATTTGGTCTCCCTCATGTTTAAT 480  
Qy 488 ATGCTGTGAGAGCTTCAACAGATGTGGGAAACATGCTGAACCTGAGTATGACCA 547  
Db 481 ATGCTGTGAGAGCTTCAACAGATGTGGGAAACATGCTGAACCTGAGTATGACCA 540  
Qy 548 ATACAGTTCACATGACAGAGATTTCTGATTTGCTATGCTATCAAGCTGAG 607

Db 541 ATACAGTTCACATGACAGAGATTTCTGATTTGCTATGCTATACAGAGCTGAG 600  
Qy 608 CTGCTTTTATTAAGAGTTCGAATGCAATTTGGCAAGGTAGCCAGAAATCAATCC 667  
Db 601 CTGCTTTTATTAAGAGTTCGAATGCAATTTGGCAAGGTAGCCAGAAATCAATCC 660  
Qy 668 GAAAGATAGCCAAAATGTCTTCTCCATCTTCAACACCTGGATCTGGTTTCACTGA 727  
Db 661 GAAAGATAGCCAAAATGTCTTCTCCATCTTCAACACCTGGATCTGGTTTCACTGA 720  
Qy 728 GCAGACAGACGGAGCTTATCTAAGGCTATGACAGAGAAACAGGGATACGTTTG 787  
Db 721 GCAGACAGACGGAGCTTATCTAAGGCTATGACAGAGAAACAGGGATACGTTTG 780  
Qy 788 TCTGAGTCTTGGTATTTAATCTCTCCATCATGTTTGAAGTATGATGCTTCACTG 847  
Db 781 TCTGAGTCTTGGTATTTAATCTCTCCATCATGTTTGAAGTATGATGCTTCACTG 840  
Qy 848 GTGTTTGTATTAAGAAATGCGGTCCAGATTGCTTGTGTAACCTTGGAAACATGGTA 907  
Db 841 GTGTTTGTATTAAGAAATGCGGTCCAGATTGCTTGTGTAACCTTGGAAACATGGTA 900  
Qy 908 CATACACAGATTCACAGTTCAGTCAACGGTGGAGAACTAGATTTGAATGAAATGA 967  
Db 901 CATACACAGATTCACAGTTCAGTCAACGGTGGAGAACTAGATTTGAATGAAATGA 960  
Qy 968 ACAAGACAGATTAATGATGCGAGTATGCTGCTATAGACTACGTCGTAATGAACTG 1027  
Db 961 ACAAGACAGATTAATGATGCGAGTATGCTGCTATAGACTACGTCGTAATGAACTG 1020  
Qy 1028 TGAAGTATTTAATGAAGATATGAAAGACAGAGATATGATGATTTTGAAGAGCT 1087  
Db 1021 TGAAGTATTTAATGAAGATATGAAAGACAGAGATATGATGATTTTGAAGAGCT 1080  
Qy 1088 ATGAGACTGCTTCATGTAAGAAATGCTCTACTCTGCGCTATGCTAACTTGGTCA 1147  
Db 1081 ATGAGACTGCTTCATGTAAGAAATGCTCTACTCTGCGCTATGCTAACTTGGTCA 1140  
Qy 1148 CTATTTTCAGTGTGGTTTAAACGCTAATATGAGTCTGCGAGTCAAGGAAATGTCGAG 1207  
Db 1141 CTATTTTCAGTGTGGTTTAAACGCTAATATGAGTCTGCGAGTCAAGGAAATGTCGAG 1200  
Qy 1208 GTACCTTACTGTTGAGATCTAGTAATGTAATGTAATGTAATGTAATGTAATGTA 1267  
Db 1201 GTACCTTACTGTTGAGATCTAGTAATGTAATGTAATGTAATGTAATGTAATGTA 1260  
Qy 1268 CCTGAACTTTCTGGGAACTGTATATAGAGACTAGCAAGCACTCATATGATGAACA 1327  
Db 1261 CCTGAACTTTCTGGGAACTGTATATAGAGACTAGCAAGCACTCATATGATGAACA 1320  
Qy 1328 CCTGTTTACTCTAATCAAGGTAGACCCCAATTAAGCAAGGTATGTCATCTCCCC 1387  
Db 1321 CCTGTTTACTCTAATCAAGGTAGACCCCAATTAAGCAAGGTATGTCATCTCCCC 1380  
Qy 1388 TTCAGATCACACACAGACAGTACCGTGGCTTGTGAATGTCATTTGAAATACATTG 1447  
Db 1381 TTCAGATCACACACAGACAGTACCGTGGCTTGTGAATGTCATTTGAAATACATTG 1440  
Qy 1448 AGGGCAGAAAGCTCTTAATGTAATATCTTTGAAGTCTCTGACAGAAAGAAAGTGCCA 1507  
Db 1441 AGGGCAGAAAGCTCTTAATGTAATATCTTTGAAGTCTCTGACAGAAAGAAAGTGCCA 1500  
Qy 1508 TTGTAGAGAGTATGAGGTCAAGGAAAGCAATATGAGAGCTATTTATTCGCTTATG 1567  
Db 1501 TTGTAGAGAGTATGAGGTCAAGGAAAGCAATATGAGAGCTATTTATTCGCTTATG 1560  
Qy 1568 AGGCTCAAAAGGTAGCTTATCTTGTGGTCAAAATATCAAGATGTGAGCTGAGAA 1627  
Db 1561 AGGCTCAAAAGGTAGCTTATCTTGTGGTCAAAATATCAAGATGTGAGCTGAGAA 1620  
Qy 1628 GCCTTGGAGGGCAGTGGAGTGTACCTCAGATGCTGTCTCTTCATATATCTATTT 1687  
Db 1621 GCCTTGGAGGGCAGTGGAGTGTACCTCAGATGCTGTCTCTTCATATATCTATTT 1680

QY 1688 ATTAACACCTCTTATATGAGAAACATCACTGCTTCACTGAGAAAGTGTATGCACTGGCAA 1747  
| | | | |  
Db 1681 ATTACAACCTCTTATATGAGAAACATCACTGCTTCACTGAGAAAGTGTATGCACTGGCAA 1740  
| | | | |  
QY 1748 AATTAGCTGACTCTATGATGCAATTTCTTGCATGCCATGATATGACACCCAGTAG 1807  
| | | | |  
Db 1741 AATTAGCTGACTCTATGATGCAATTTCTTGCATGCCATGATATGACACCCAGTAG 1800  
| | | | |  
QY 1808 GGAAGCAGACTCAAGCTTCAGAGAGAGAAAGCAAGAGTGGCAATTGCAAGAGCCA 1867  
| | | | |  
Db 1801 GGAAGCAGACTCAAGCTTCAGAGAGAGAAAGCAAGAGTGGCAATTGCAAGAGCCA 1860  
| | | | |  
QY 1868 TTTTGAAGACCCCCCACTACTCTATGATGAGACTTCTTCACTTGAATTGATTA 1927  
| | | | |  
Db 1861 TTTTGAAGACCCCCCACTACTCTATGATGAGACTTCTTCACTTGAATTGATTA 1920  
| | | | |  
QY 1928 CTGAAGAGACTATTTCTTGTCGCCATGAAAGATGTGTCAACACAGAACTTTTCA 1967  
| | | | |  
Db 1921 CTGAAGAGACTATTTCTTGTCGCCATGAAAGATGTGTCAACACAGAACTTTTCA 1960  
| | | | |  
QY 1988 TTGCACACAGATTTGTCAACAGTGTGATGAGATGAAATCATTTGTGTGATCAGGGTA 2047  
| | | | |  
Db 1981 TTGCACACAGATTTGTCAACAGTGTGATGAGATGAAATCATTTGTGTGATCAGGGTA 2040  
| | | | |  
QY 2048 AGGTAGCCGAACTGTGTACCCACCATGTTGCTTGTAACTTCATATGATCTATTCA 2107  
| | | | |  
Db 2041 AGGTAGCCGAACTGTGTACCCACCATGTTGCTTGTAACTTCATATGATCTATTCA 2100  
| | | | |  
QY 2108 AATATGTGCATACACAGACGCCGTGTGCAGAAACATGATTAACCCCAATGGAGCAA 2167  
| | | | |  
Db 2101 AATATGTGCATACACAGACGCCGTGTGCAGAAACATGATTAACCCCAATGGAGCAA 2160  
| | | | |  
QY 2168 AGAAGAAATATATATCAAGAGAGAGAGAAAGAAAGAACTCAAGAGAAATTTGCATA 2227  
| | | | |  
Db 2161 AGAAGAAATATATATCAAGAGAGAGAGAAAGAAAGAACTCAAGAGAAATTTGCATA 2220  
| | | | |  
QY 2228 GTGTGAAGGCTGTGAAACTGTTCGTCTAGTCACATTAAGACATTTTCTTTTGT 2286  
| | | | |  
Db 2221 GTGTGAAGGCTGTGAAACTGTTCGTCTAGTCACATTAAGACATTTTCTTTTGT 2278  
| | | | |

RESULT 6  
AR100662 2407 bp DNA linear PAT 14-FEB-2001  
LOCUS Sequence 2 from patent US 6080842.  
DEFINITION AR100662  
ACCESSION AR100662  
VERSION AR100662.1 GI:12811110  
KEYWORDS

SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2407)  
AUTHORS Hillman,J.L., Shah,P. and Corley,N.C.  
TITLE Human ATP binding-cassette transport protein  
JOURNAL Patent: US 6080842-A 2 27-JUN-2000;  
FEATURES  
source location/Qualifiers  
1..2407  
/organism="unknown"

BASE COUNT 726 a 477 c 570 g 629 t 5 others  
ORIGIN

Query Match 95.9%; Score 2248; DB 6; Length 2407;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 2265; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 8 TGCTCGGATGATCTTCTTGCGCTGAGCGCGCGCGCTGCTTTGCAAAAGCCGCGC 67  
| | | | |  
Db 1 TGCTCGGATGATCTTCTTGCGCTGAGCGCGCGCGCTGCTTTGCAAAAGCCGCGC 60  
| | | | |  
QY 68 ACTCGGATGATCTTCTTGCGCTGAGCGCGCGCTGCTTTGCAAAAGCCGCGC 127  
| | | | |  
Db 61 ACTCGGATGATCTTCTTGCGCTGAGCGCGCGCTGCTTTGCAAAAGCCGCGC 120  
| | | | |

QY 128 CACATCAACTCGGCGCTTTGGGAACCGCTCGACCTACAGATTTCCAGATCTTAAAAA 187  
| | | | |  
Db 121 CACATCAACTCGGCGCTTTGGGAACCGCTCGACCTACAGATTTCCAGATCTTAAAAA 180  
| | | | |  
QY 188 GTATCAATGCGCAGAGATTTGGGAAAGGCAATTCAGACAGTTCTTATGATGCTGCAAGG 247  
| | | | |  
Db 181 GTATCAATGCGCAGAGATTTGGGAAAGGCAATTCAGACAGTTCTTATGATGCTGCAAGG 240  
| | | | |  
QY 248 CTCCTCCAGATGCGCAGCTGATGAAAGAGGACATGTTGGCATGGGTCAAGCAGAGAG 307  
| | | | |  
Db 241 CTCCTCCAGATGCGCAGCTGATGAAAGAGGACATGTTGGCATGGGTCAAGCAGAGAG 300  
| | | | |  
QY 308 GACTCCACAGACCCCAAAAGAGGTTAAAGATGTTGATCTCGGAAATCATAAAG 367  
| | | | |  
Db 301 GACTCCACAGACCCCAAAAGAGGTTAAAGATGTTGATCTCGGAAATCATAAAG 360  
| | | | |  
QY 368 CAATGCTTTCTTATGTGTGCGCCAAAGACAGGCCAGATCTTACAGACTAGAGTTGCCATTT 427  
| | | | |  
Db 361 CAATGCTTTCTTATGTGTGCGCCAAAGACAGGCCAGATCTTACAGACTAGAGTTGCCATTT 420  
| | | | |  
QY 428 CGCTGGGATTTTGGGTGGTGCAGAAAGGCCATGAATATGTGGTTCCCTTCATGTTTAAAT 487  
| | | | |  
Db 421 CGCTGGGATTTTGGGTGGTGCAGAAAGGCCATGAATATGTGGTTCCCTTCATGTTTAAAT 480  
| | | | |  
QY 488 ATGCTGAGACAGCTCAACCAAGATGTCGGGAAACATGCTGAACCTGATGATGACCAA 547  
| | | | |  
Db 481 ATGCTGAGACAGCTCAACCAAGATGTCGGGAAACATGCTGAACCTGATGATGACCAA 540  
| | | | |  
QY 548 ATACAGTTGCAACCATGCAACAGCAGTTCTGATGCTATGCTGATATCAAGAGCTGGAG 607  
| | | | |  
Db 541 ATACAGTTGCAACCATGCAACAGCAGTTCTGATGCTATGCTGATATCAAGAGCTGGAG 600  
| | | | |  
QY 608 CTCCTTTTAAACGAAGTTCCGAATGCGAGTATTTGGCAAGATAGCCCAATTCATTC 667  
| | | | |  
Db 601 CTCCTTTTAAACGAAGTTCCGAATGCGAGTATTTGGCAAGATAGCCCAATTCATTC 660  
| | | | |  
QY 668 GAAGAAATGCGCAAAATGCTTCTTCATCTTCAACCTGATCTGGGTTTCACTGA 727  
| | | | |  
Db 661 GAAGAAATGCGCAAAATGCTTCTTCATCTTCAACCTGATCTGGGTTTCACTGA 720  
| | | | |  
QY 728 GCAGACAGACGGAGCTTTATCTAAGGCTATTTGACAGAGAACAGGGGTATCAGTTTGG 787  
| | | | |  
Db 721 GCAGACAGACGGAGCTTTATCTAAGGCTATTTGACAGAGAACAGGGGTATCAGTTTGG 780  
| | | | |  
QY 788 TCCGTAGTGTGTTGATTTAATCTTCTCCATCATGTTTGAAGATGCTGTCAAGT 847  
| | | | |  
Db 781 TCCGTAGTGTGTTGATTTAATCTTCTCCATCATGTTTGAAGATGCTGTCAAGT 840  
| | | | |  
QY 848 GTGTTTGTATTAACAATGCGGTGCGCAGTTTGTGTTGAACTTGGAACTTGGTA 907  
| | | | |  
Db 841 GTGTTTGTATTAACAATGCGGTGCGCAGTTTGTGTTGAACTTGGAACTTGGTA 900  
| | | | |  
QY 908 CATACACAGATTCACAGTTGCGAGTCAACCGGTGAGAACTAGATTTAGAAATGA 967  
| | | | |  
Db 901 CATACACAGATTCACAGTTGCGAGTCAACCGGTGAGAACTAGATTTAGAAATGA 960  
| | | | |  
QY 968 ACAAGCAGATTAATGATGAGGATATGCTGCTATTAACATCACTGCTGAATTAGAAACTG 1027  
| | | | |  
Db 961 ACAAGCAGATTAATGATGAGGATATGCTGCTATTAACATCACTGCTGAATTAGAAACTG 1020  
| | | | |  
QY 1028 TGAAGTATTTAATATGAAAGATATGAAGCAGAGATATGATGATTTTGAAGCGT 1087  
| | | | |  
Db 1021 TGAAGTATTTAATATGAAAGATATGAAGCAGAGATATGATGATTTTGAAGCGT 1080  
| | | | |  
QY 1088 ATGAGACTGCTTATGAAAGATACCTTACTGTGCTATGCTGAACCTTGTGCAAAAGTG 1147  
| | | | |  
Db 1081 ATGAGACTGCTTATGAAAGATACCTTACTGTGCTATGCTGAACCTTGTGCAAAAGTG 1140  
| | | | |  
QY 1148 CTATTTTCAAGTGTGCTTTTAAACAGCTATAATGCTGCTGCGCACTCAGGGAATTTGGCAG 1207  
| | | | |  
Db 1141 CTATTTTCAAGTGTGCTTTTAAACAGCTATAATGCTGCTGCGCACTCAGGGAATTTGGCAG 1200  
| | | | |

QY 1208 GTACCTTAAGTGTGAGATCTAGTAATGTGAATGAGACTGCTTTTTCAGCTTTTCATTAC 1267  
 |||||  
 DB 1201 GTACCTTAAGTGTGAGATCTAGTAATGTGAATGAGACTGCTTTTTCAGCTTTTCATTAC 1260  
 |||||  
 QY 1268 CCTTGAACCTTTCTGGGAAGTGTATATAGAGACTAGACACTCATAGATATGAACA 1327  
 |||||  
 DB 1261 CCTTGAACCTTTCTGGGAAGTGTATATAGAGACTAGACACTCATAGATATGAACA 1320  
 |||||  
 QY 1328 CCTTGTAACTCTACTCAAGTAGAGACCCCAATTTAAAGACAAAGTAGATGATCCTCCC 1387  
 |||||  
 DB 1321 CCTTGTAACTCTACTCAAGTAGAGACCCCAATTTAAAGACAAAGTAGATGATCCTCCC 1380  
 |||||  
 QY 1388 TTCAATGATCACACAGACAGCTACCGTGGCTTTGATATGTGATGATTTTGAATACATTG 1447  
 |||||  
 DB 1381 TTCAATGATCACACAGACAGCTACCGTGGCTTTGATATGTGATGATTTTGAATACATTG 1440  
 |||||  
 QY 1448 AGGGCCAGAAATCTCTTATGTGAATATCTTTGAAAGTCCCTCAGAGAAAGAAAGTGGCCA 1507  
 |||||  
 DB 1441 AGGGCCAGAAATCTCTTATGTGAATATCTTTGAAAGTCCCTCAGAGAAAGAAAGTGGCCA 1500  
 |||||  
 QY 1508 TTGTAGAGAGTGTGAGGAGTCAAGGAAAGACAAATAGTGAAGGCTATTTATTTGCTTATG 1567  
 |||||  
 DB 1501 TTGTAGAGAGTGTGAGGAGTCAAGGAAAGACAAATAGTGAAGGCTATTTATTTGCTTATG 1560  
 |||||  
 QY 1568 AGCTCAAAAGGGTAGCATTTATCTTGTCTGTCAAAATATACAGATGTGAGCTTGGAAA 1627  
 |||||  
 DB 1561 AGCTCAAAAGGGTAGCATTTATCTTGTCTGTCAAAATATACAGATGTGAGCTTGGAAA 1620  
 |||||  
 QY 1628 GCGTTCGAGAGGAGTGGAGAGTGTATACCTCAGAGATGCTGCTCTTCCATATCTATTT 1687  
 |||||  
 DB 1621 GCGTTCGAGAGGAGTGGAGAGTGTATACCTCAGAGATGCTGCTCTTCCATATCTATTT 1680  
 |||||  
 QY 1688 ATTACAACTCTTATATGGAACATCAAGTGTCACTGAGAGAGTGTGAGTGGCA 1747  
 |||||  
 DB 1681 ATTACAACTCTTATATGGAACATCAAGTGTCACTGAGAGAGTGTGAGTGGCA 1740  
 |||||  
 QY 1748 AATTAGCTGAGCTTCAATGATGCAATTTCTTGAATGCAATGATGATGATGATGATGATG 1807  
 |||||  
 DB 1741 AATTAGCTGAGCTTCAATGATGCAATTTCTTGAATGCAATGATGATGATGATGATGATG 1800  
 |||||  
 QY 1808 GGGAGCAGAGATCTCAAGCTTTCAAGAGAGAAAGAAAGAGTGAATGCAATGCAAGGCA 1867  
 |||||  
 DB 1801 GGGAGCAGAGATCTCAAGCTTTCAAGAGAGAAAGAAAGAGTGAATGCAATGCAAGGCA 1860  
 |||||  
 QY 1868 TTTTGAAGAGCCCGCAGTCAATCTATGATGAGAGCTACTTCAATGATGATGATGATGATG 1927  
 |||||  
 DB 1861 TTTTGAAGAGCCCGCAGTCAATCTATGATGAGAGCTACTTCAATGATGATGATGATGATG 1920  
 |||||  
 QY 1928 CTGAAGAGCTATTTCTTGTGTCATGAGAGATGTGTCAACAGAGAACTTCTATTTTCA 1987  
 |||||  
 DB 1921 CTGAAGAGCTATTTCTTGTGTCATGAGAGATGTGTCAACAGAGAACTTCTATTTTCA 1980  
 |||||  
 QY 1988 TTGCAACAGATTTGTCAACAGAGTGTGATGCAATGATGATGATGATGATGATGATGATG 2047  
 |||||  
 DB 1981 TTGCAACAGATTTGTCAACAGAGTGTGATGCAATGATGATGATGATGATGATGATGATG 2040  
 |||||  
 QY 2048 AGGTAGCCGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2107  
 |||||  
 DB 2041 AGGTAGCCGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100  
 |||||  
 QY 2108 AAATGTGCAATCAGAGAGAGCCGTGTGAGAGACCATGATTAACCCCAATGGAGAGCA 2167  
 |||||  
 DB 2101 AAATGTGCAATCAGAGAGAGCCGTGTGAGAGACCATGATTAACCCCAATGGAGAGCA 2160  
 |||||  
 QY 2168 AGAAAGAAATATATTCAG 2227  
 |||||  
 DB 2161 AGAAAGAAATATATTCAG 2220  
 |||||  
 QY 2228 GTGTAGAGAGCTGTGAGAGAGCTGTGTGTGATGATGATGATGATGATGATGATGATGATG 2286  
 |||||  
 DB 2221 GTGTAGAGAGCTGTGAGAGAGCTGTGTGTGATGATGATGATGATGATGATGATGATGATG 2278  
 |||||

RESULT 7  
 AK001418 2193 bp mRNA linear PRI 01-AUG-2002  
 LOCUS Homo sapiens cDNA FLJ10556 fis, clone NT2RP2002479, highly similar  
 DEFINITION to Homo sapiens mRNA for ABC transporter 7 protein.  
 ACCESSION AK001418  
 VERSION AK001418.1 GI:7022662  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens testocarcinoma cell line:NT2 cDNA to mRNA,  
 clone 11b:NT2RP2 clone:NT2RP2002479.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
 Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,  
 Tanase,T., Nomura,Y., Togiyu,S., Komai,F., Hara,R., Takeuchi,K.,  
 Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,  
 Wakamatsu,A., Nakamura,Y., Nagaiari,K., Masuno,Y. and Ohshima,A.  
 NEDO human cDNA sequencing project  
 TITLE NEDO human cDNA sequencing project  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2193)  
 AUTHORS Isegai,T. and Otsuki,T.  
 JOURNAL Direct Submission  
 TITLE Submitted (16-FEB-2000) Takao Isegai, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 COMMENT NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan. cDNA full insert  
 sequencing: Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection:  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.  
 FEATURES  
 source  
 1. 2193  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP2002479"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_11b="NT2RP2"  
 /note="Cloning vector: pME18SFL3-mRNA from NT2 neuronal  
 precursor cells after 2-weeks retinoic acid (RA)  
 induction."  
 BASE COUNT 621 a 451 c 536 g 585 t  
 ORIGIN  
 Query Match 91.7%; Score 2150.4; DB 9; Length 2193;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 2164; Conservative 0; Mismatches 1; Indels 3; Gaps 1;  
 QY 1 ATGGCGCTGCTCGGATGATGATTTCTTGGGCGGCGCGGCGGCGCTTTGGAAAG 60  
 |||||  
 DB 26 ATGGCGCTGCTCGGATGATGATTTCTTGGGCGGCGGCGGCGGCGCTTTGGAAAG 85  
 |||||  
 QY 61 GCGCGCACTCGGATCTGATCGGCGCTTAACTCTGTGTAAGCGGCTCAGTCCGAG 120  
 |||||  
 DB 86 GCGCGCACTCGGATCTGATCGGCGCTTAACTCTGTGTAAGCGGCTCAGTCCGAG 145  
 |||||  
 QY 121 TGAAGCCACATCACTGGGCGCTTGGGAACGCTCGAGCTTAC--CAGATTCAGAG 177  
 |||||  
 DB 146 TGAAGCCACATCACTGGGCGCTTGGGAACGCTCGAGCTTAC--CAGATTCAGAG 205  
 |||||  
 QY 178 TCATTAAAAAGTATCACTGGCAGAGATTTGGGAAAGGCAATTCAAGACAGTTCTTAGAT 237  
 |||||  
 DB 206 TCATTAAAAAGTATCACTGGCAGAGATTTGGGAAAGGCAATTCAAGACAGTTCTTAGAT 265  
 |||||  
 QY 238 GGTGAAAGGCTTCGAGGTATGGCACTGATAGAAAAGAGACATGTTGCAATGATGATGATGATG 297  
 |||||  
 DB 266 GGTGAAAGGCTTCGAGGTATGGCACTGATAGAAAAGAGACATGTTGCAATGATGATGATGATG 325  
 |||||  
 QY 298 GCAGAGAGAGACTCCACAGACCAAGAGAGGTTAAAGATGTTGATATCTCGAAA 357  
 |||||







PUBMED 9143506  
REFERENCE 2 (bases 1 to 2684)  
AUTHORS Savary,S., Denizot,F., Matrei,M.G. and Chimini,G.  
TITLE Direct Submission  
JOURNAL Submitted (26-DEC-1995) Stephane Savary, Centre d'Immunologie de  
Marseille Luminy, Parc Scientifique et Technologique de Luminy,  
Case 906, Marseille Cedex 9, 13288, France

FEATURES  
source  
1..2684  
/organism="Mus musculus"  
/strain="DBA/2"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/map="XC-D"  
/clone="lambda ABC7-3b"  
/cell\_line="P388D1"  
/cell\_type="monocyte/macrophage"  
1..2085  
/codon\_start=1  
/product="ABC transporter-7"  
/protein\_id="AAC5152.1"  
/db\_xref="GI:1167982"  
/translation="ESLRNTTQQRKGNKNSROLLDATALKLTWPLEIKRTGHHAG  
GLHDPKSLGKDVDTKTIKAMLSYVPEDEPDLRARAISLGLGAKKAMIVPEM  
FKYAVDSLQMSGMNLSDAPNTVAATAVLIGVSRGAAPFNEVRNAVGVKA  
ONSIRIRIKNPFILHNLDFHLSRQTGALSKAIDRTGRISEVLNPLPIVF  
EMMLVSVLYKCGAOFALVTIGTATFAVTAFTVTRRTFRLEMKKADNAGNAI  
DSLNTETVKKYFNNEKYEAPRQDGIKRTETASIKSTSTAMLNPKADIPSVGLTAI  
MYLASQIVAAALVTGDLVWNGLLFOLSLNPLNGVYRRTKQALIDMTLFTLLKY  
DTRIKDKVMAPLDITPQTATVADNVFETIEQKLVNVSFVPPAKKVAITVLSGS  
SGKSTVRLFRFEPKQSLYLQGNLQDVLSLRVAVVPODVALFNNTIYML  
LYGINSNPBEVYVAKLAGLHDAILRMPHYDYOVRGKLSGEKQRYAIVRAIL  
KNPVLIDENATSSLSITETILGAMNDVHKRSTFIARLSVVDADIEIYLSOG  
KVAREGTHYGLANSSSIYSEMHMTQSNRVQNSLGDAKKESLSKEBEKKLQBEI  
VNSYKGCNCSC"

CDS  
BASE COUNT 818 a 502 c 581 g 782 t 1 others  
ORIGIN

Query Match 78.2%; Score 1833; DB 10; Length 2684;  
Best Local Similarity 91.4%; Pred. No. 0;  
Matches 1941; Conservative 1; Mismatches 181; Indels 0; Gaps 0;

Qy 175 GAGTCATTAAAAAGTATCATGACATGGACAGAGATTGGAAAAAGCAATTCCAGACAGTTCTTA 234  
Db 1 GAATCATTTAAAGAAATACCTACACAGACAGATGGGAAAAAGACATCCACAGATTACTA 60  
Qy 225 GATGCTGCAAAAGCTCTCCAGGTATGGCCACTGATGAAAAAGACATGTTGGCATGTT 294  
Db 61 GATGCTACAAAGCTCTCCAGCATGGCCATTGATGAAAAAGACATGTTGGCATGG 120  
Qy 295 CATGCAGAGAGAGACTCCACACAGACCCAAAGAAAGGTTAAAGATGTGATCTGG 354  
Db 121 CACGCGAGGAGAGACTCCACACAGACCCAAAGAAAGGTTAAAGATGTGATCTGA 180  
Qy 355 AAAATCATAAAAAGCAATGCTTTCTTATGTGTGCCCCAAAGACAGGCGAGATCTACAGCT 414  
Db 181 AAAATCATAAAGCATGCTTTCTTATGTGTGCCCCAAAGACAGATCTACAGAGCC 240  
Qy 415 AGAGTTGCCATTTCCGTGGGATTTTGGGTGGTGCAAAGGCCATGAAATTTGTGTTCCC 474  
Db 241 AGAGTTGCCATCTCCCTGGGATTTTGGGTGGTGCAAAGGCCATGAAATTTGTGTTCC 300  
Qy 475 TTTCATGTTTAAATATGCTGTAGACAGCCCTCAACCGATGCGGAAACATGCTGAACCTG 534  
Db 301 TTTCATGTTTAAATATGCTGTAGACAGCCCTCAACCGATGCGGAAACATGCTGAACCTG 360  
Qy 535 AGTGTATGACCAAAATACAGTTGCAACCATGGAACAGCAGATTCTGATTTGGCTATGGTGA 594  
Db 361 AGTGTATGACCAAAATACAGTTGCAACCATGGAACAGCAGATTCTGATTTGGCTATGGTGA 420  
Qy 595 TCAAGAGCTGAGAGCTGCTTTTAAACGAAGTTCCAAATGACGATTTTGGCAGAGTAGCC 654  
Db 421 TCAAGAGCTGAGAGCTGCTTTTAAACGAAGTTCCAAATGACGATTTTGGCAGAGTAGCA 480

Qy 655 CAGATTCATATCCAGAGATAGCCAAAAATGCTCTTCTCATCTTCAACACTGGATCTG 714  
Db 481 CAATATCAATACAGAGATAGCCAAAAATGATTTCTTCATCTTCCACACTGGATCTG 540  
Qy 715 GGTTTTCACTGACGACAGACAGCGGAGCTTTATCTAAGGCTATTTGACAGAGAAACAAG 774  
Db 541 GGTTTTCACTGACGACAGACAGCGGAGCTTTATCTAAGGCTATTTGACAGAGAAACAAG 600  
Qy 775 GGATACGTTTGTGCTGAGAGCTTTGGTATTTAATGCTCTTCCATCATATTTGAAGTG 834  
Db 601 GGATACGTTTGTGCTGAGAGCTTTGGTATTTAATGCTCTTCCATCATATTTGAAGTG 660  
Qy 835 ATGCTTGTCACTGGTGTGTTTGTATTAACAATCCGTCGCCAGTTTGGTAAACCTT 894  
Db 661 ATGCTTGTCACTAGTGTGTTTGTATTAACAATGGTGCCAGTTTGGCATTTGTAACCTTA 720  
Qy 895 GGAACACTTGGTATACACACAGCATTTCACTGACAGTACACGCTGAGAACTAGATTT 954  
Db 721 GGAACACTTGGTATATACAGCATTTCACTGACAGTACACGCTGAGAACTAGATTT 780  
Qy 955 AGAATAGAAATGAACAAAGCAGATATGATGACGTAATGCTGCTATAGACTCATGCTG 1034  
Db 781 AGAATAGAAATGAACAAAGCAGATATGATGACGTAATGCTGCTATAGACTCATGCTG 840  
Qy 1015 AATTATGAATCTGAAATATTTTAAATGAAAGATATGAAAGACAGAGATATGATGGA 1074  
Db 841 AATTATGAATCTGAAATATTTTAAATGAAAGATATGAAAGACAGAGATATGATGGA 900  
Qy 1075 TTTTGAAGACGTATGAGACTGCTTCACTGAAAGATCTTCACTCTGGCTATGCTGAA 1134  
Db 901 TTTTGAAGACGTATGAGACTGCTTCACTGAAAGATCTTCACTCTGGCTATGCTGAA 960  
Qy 1135 TTTTGAAGACGTATGAGACTGCTTCACTGAAAGATCTTCACTCTGGCTATGCTGAA 1194  
Db 961 TTTTGAAGACGTATGAGACTGCTTCACTGAAAGATCTTCACTCTGGCTATGCTGAA 1020  
Qy 1195 GGAATTTGGCAGGATGACCTTCTACTGTTGAGATCTAGTAATGTAATGATGATCTTT 1254  
Db 1021 GGAATTTGGCAGGATGACCTTCTACTGTTGAGATCTAGTAATGTAATGATGATCTTT 1080  
Qy 1255 CAGCTTTCATTAACCTCGAACTTTCTGGGAACTGATATGAGAGACTGACAGACACTC 1314  
Db 1081 CAGCTTTCATTAACCTCGAACTTTCTGGGAACTGATATGAGAGACTGACAGACACTC 1140  
Qy 1315 ATGATATGAACACTGTTTACTCTACTCAAGGTAGCACCCAAATTTAAAGACAAAGTG 1374  
Db 1141 ATGATATGAACACTGTTTACTCTACTCAAGGTAGCACCCAAATTTAAAGACAAAGTG 1200  
Qy 1375 ATGCAATCTCCCTTTCAGATCACACACAGACAGTACCGTGCTTGAATATGTCAT 1434  
Db 1201 ATGCAATCTCCCTTTCAGATCACACACAGACAGTACCGTGCTTGAATATGTCAT 1280  
Qy 1435 TTTGATATCATTTGAGGCGCAGAAAGTCTTATGAGAAATTCCTTTGAAGTCTTCAGGA 1494  
Db 1261 TTTGATATCATTTGAGGCGCAGAAAGTCTTATGAGAAATTCCTTTGAAGTCTTCAGGA 1320  
Qy 1495 AAGAAAGTGGCATTTGATAGAGATAGTGGGTCAAGGAAACCAATATGACGCTATTA 1554  
Db 1321 AAGAAAGTGGCATTTGATAGAGATAGTGGGTCAAGGAAACCAATATGACGCTATTA 1380  
Qy 1555 TTTGCTTCTTATGAGAGCTCAAAAGGATGACATTTATCTTGGTGTCAAAATATCAAGAT 1614  
Db 1381 TTTGCTTCTTATGAGAGCTCAAAAGGATGACATTTATCTTGGTGTCAAAATATCAAGAT 1440  
Qy 1615 GTGAGCTTGAAAGCTTTTGGAGGCGCAGTGGAGTGTATCTCAGATGCTGTCTTTC 1674  
Db 1441 GTGAGCTTGAAAGCTTTTGGAGGCGCAGTGGAGTGTATCTCAGATGCTGTCTTTC 1500  
Qy 1675 CATATATATTTATTTTCAACCTCTTATATGAGAAACATCACTGCTTCACTGAGAGAGTG 1734  
Db 1501 CATATATATTTATTTTCAACCTCTTATATGAGAAACATCACTGCTTCACTGAGAGAGTA 1560

QY 1735 TATGAGTGGCAAAATTAAGTGGACCTGATGATGCAATTTCTTGGAATGCCATGATAT 1794  
 |||||  
 DB 1561 TATGAGTGGCAAAATTAAGTGGACCTGATGATGCAATTTCTTGGAATGCCATGATAT 1620  
 |||||  
 QY 1795 GACACCAAGAGGGAACGAGAGCTCAAGCTTCAGAGAGGAAAGCAAGAGTGA 1854  
 |||||  
 DB 1621 GACACCAAGAGGGAACGAGAGCTCAAGCTTCAGAGAGGAAAGCAAGAGTGA 1660  
 |||||  
 QY 1855 ATTGCAAGAGCTTTTGAAGAGACCCCACTGATCTATCTATGATGAAGTACTTCAATG 1914  
 |||||  
 DB 1681 ATTGCAAGAGCTTTTGAAGAGATCCCACTATATCTATGATGAAGTACTTCAATG 1740  
 |||||  
 QY 1915 TTGATTCGATTACTGAAGAGACTATTTCTTGTCATGAAGATGTGTCAACACAGA 1974  
 |||||  
 DB 1741 TTGATTCGATTACTGAAGAGACTATTTCTTGTCATGAAGATGTGTCAACACAGA 1800  
 |||||  
 QY 1975 ACTTCTATTTTCAATGCAACAGATGTGCAACAGTGTGTGAAGATGAATGATGTC 2034  
 |||||  
 DB 1801 ACTTCTATTTTCAATGCAACAGATGTGCAACAGTGTGTGAAGATGAATGATGTC 1860  
 |||||  
 QY 2035 TTGATTCGATTACTGAAGAGACTATTTCTTGTCATGAAGATGTGTCAACACAGA 2094  
 |||||  
 DB 1861 CTGAGCCAGGGAAGATGAGTGAAGTGTGACCACTATGTCTGTCTGCTTAAGTAC 1920  
 |||||  
 QY 2095 AGTATCTATTCAGAAATGTGCAATACAGAGACGCGTGTGCAAGAACATGATTAACCC 2154  
 |||||  
 DB 1921 AGTATCTATTCAGAAATGTGCAATACAGAGACGCGTGTGCAAGAACATGATTAACCC 1980  
 |||||  
 QY 2155 AAAATGGAGCAAGAAAGAAATATATCCAAAGAGAGGAAAGAAAGAACTCAAGAA 2214  
 |||||  
 DB 1981 GGAATGGAGTGAAGAAAGAAAGATCTCCAAAGAGAGGAAAGAAAGAACTCAAGAA 2040  
 |||||  
 QY 2215 GAAATTTGCAATAGTGTGAAGAGCTGTGGAAGAACTGTGCTAAGTCACTAAGACATT 2274  
 |||||  
 DB 2041 GAAATTTGCAATAGTGTGAAGAGCTGTGGAAGAACTGTGCTAAGTCACTAAGACATT 2100  
 |||||  
 QY 2275 TTCTTTTGTGTTGTTGACT 2297  
 |||||  
 DB 2101 TCTATCTCTCTTTTGTGTTCT 2123  
 |||||

RESULT 9  
 AY051556 2979 bp mRNA linear INV 27-AUG-2001  
 LOCUS Drosophila melanogaster GH20617 full length cDNA.  
 DEFINITION AY051556 GI:15291422  
 VERSION AY051556.1 GI:15291422  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Drosophila melanogaster.  
 Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 2979)  
 REFERENCE  
 AUTHORS  
 Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,  
 Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R.,  
 Gonzalez, M., Guarin, H., Li, P., Liao, G., Miranda, A., Mungall, C. J.,  
 Nuno, J., Pacleb, J., Parag, V., Park, S., Phoumenavong, S., Wan, K.,  
 Yu, C., Lewis, S. E., Rubin, G. M. and Celniker, S.  
 TITLE  
 JOURNAL  
 Submitted (10-AUG-2001) Berkeley Drosophila Genome Project,  
 Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720, USA  
 COMMENT  
 Sequence submitted by:  
 Lawrence Berkeley National Laboratory  
 Berkeley Drosophila Genome Project  
 Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to  
 sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
 Science 2000). The sequence has been subjected to integrity checks  
 for sequence accuracy. Presence of a polyA tail and contiguity  
 within 100 kb in the genome. Thus we believe the sequence to  
 reflect accurately this particular cDNA clone. However, there are

artifacts associated with the generation of cDNA clones that may  
 have not been detected in our initial analyses such as internal  
 priming, priming from contaminating genomic DNA, retained introns  
 due to reverse transcription of unspliced precursor RNAs, and  
 reverse transcriptase errors that result in single base changes.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our web site  
 (<http://fruitfly.berkeley.edu>) or send email to  
 cdna@fruitfly.berkeley.edu.

## FEATURES

source  
 location/Qualifiers  
 1..2979  
 /organism="Drosophila melanogaster"  
 /strain="Y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /map="62A9-62A9"

## gene

1..2979  
 /gene="CG7955"  
 /note="alignment with genomic scaffold AB003472"  
 /db\_xref="FLYBASE:FBgn0035244"  
 717..2537  
 /gene="CG7955"  
 /note="Longest ORF"  
 /codon\_start=1  
 /product="GH20617p"  
 /protein\_id="AAK92980.1"  
 /db\_xref="GI:15291423"

## CDS

/translation="MIRAMAYIMPEDDILVRKRVISIGLAGSKLITVCPFLFG  
 AYDPTMTLMMDTPADPAVLSATATLMLGYIAASAAGNELRAVPAKRAHSIRKIA  
 SVFPLHLNLDIAFLNKOTGALSKTIDSGSGINVLASAVFIVPTIFELAVSSI  
 LKVKCGIAPAGSMGCVGYAAVYTLSTVDMRTQFRFQNAQENEAQKAVDILINET  
 VYFENNEKEACGVNEVLKRYEASLSTALNFCQNAIFSSLSLIMVLAKEI  
 AQGNMTVGLVWVNLIPOLSIPLGIFGVYEBVRALLDMNAFLMVDSSICOTRA  
 NAQPLEVDPTNNSISRRNYSREYREPKRIFRLSTPIPKGRVAVYMGSGSSWVR  
 ILRFREPSGVYLIGGODISAVDESLKVIYVPODSVLPHNTIENHIVGNLSKS  
 HAEVQNAARMDLHDSIMSWPGQYSTQVGERGLKLSGEKQVALARAILIKPTPLIF  
 DEATSLDISITENHILQALTRATSGRTSICIRHSTVDADEILVLENGVRGERGTH  
 SELIRONGYARLMEOTOTQOFPDSREINEVAAKTRGVA"

BASE COUNT 722 a 750 c 760 g 727 t  
 ORIGIN

Query Match 24.1%; Score 564.8; DB 3; Length 2979;  
 Best Local Similarity 58.5%; Pred. No. 9,8e-128;  
 Matches 1011; Conservative 0; Mismatches 702; Indels 15; Gaps 1;

QY 354 GAAATTCATTAAGCAATGCTTTCTTATGTGTGGCCCAAGAGAGCCCAATCTAGAC 413  
 |||||  
 DB 713 GGAATATGCTGCGGCGCATGATGCGCTTACATCTGCGCAAGAGATCCACTGTCGAAA 772  
 |||||  
 QY 414 TAGAGTTGCATTTGGGTGGATTTTGGGTGGTGAAGAGCCATGAATATTGTTCC 473  
 |||||  
 DB 773 GCGGGTGGATTTTCCCTGTGGGTCTGTGGCTGCTCAAGCTGTAACCTGTGTGCC 832  
 |||||  
 QY 474 CTTCATGTTTAAATATGCTGTAGACAGCTTCACAGATGTCGGGAATGCTGAACCT 533  
 |||||  
 DB 833 CTTCGTTTCAAGAGAGCTGTGACA-----CGATGACAACGCTGAACAT 877  
 |||||  
 QY 534 GAGTATGACCAAAATACAGTTGCAACCATGCAACAGCAGTTCTGATTGGCTATGGTGT 593  
 |||||  
 DB 878 GGAACACTGCCCCCGATGCGGTCTCTCCGAGCCACCGCATATGTTGGGAATGTCAT 937  
 |||||  
 QY 594 ATGAAGAGCTGAGAGCTGCTTTTAAAGAGTTCGAATGAGATATTGGCAAGTAGC 653  
 |||||  
 DB 938 TGTATAGCCAGTGCAGCGGCGCTTTAATGAGTTGCGGAATGCAAGTGTTCGAAAGTGGC 997  
 |||||  
 QY 654 CCAAGATTCATTCGAAGATAGCAAAATGCTTTCTTCATCTTCACAACTGATCT 713  
 |||||  
 DB 998 CCAACCACTGATTCGAAGATAGCAATGCTTCCTGCACTGCAACCACTGATCT 1057  
 |||||  
 QY 714 GGGTTTACCTGAGACAGACAGCGGAGCTTTATCTAAGCTTATGACAGAGAAACAG 773  
 |||||  
 DB 1058 GGCCTTCCACTGAACAACTGAGAGCACTGTCAAAAGACCATGATGAGAGATCGAG 1117  
 |||||

|    |      |   |      |
|----|------|---|------|
| Oy | 774  | GGGATACAGTTTGTCTGTAGAGCTTTGGATTTAACTCTTCCACATCAGTTTGAAGT        | 833  |
| Db | 1118 | GGGAGATTAACTTTGTGTCTCCGCGCAGTGGCTTCAACATTTGGCCCATCTATTTTGAGCT   | 1177 |
| Oy | 834  | GATCGTTGTCAAGTGTGTTTTGTATTACAATATGCGGTGCCAGTTTGGCTTGTAAACCT     | 893  |
| Db | 1178 | GGCCCTGTGTCCAGTATCCTGGGAATGAAGTGTGCGCTTGGCTTCCGTGTGTCAAGAT      | 1237 |
| Oy | 894  | TGGAACCTTGATACATACACAGCATTTCAACAGTTGCATCACAGGTGGAACCTAGATT      | 953  |
| Db | 1238 | GGGCTGCGTTGGCAATATACCTGCTCAACTGTAGGTGACCCAGTGGCCACCCAGTT        | 1297 |
| Oy | 954  | TAGAAATGAAAATGACAAACAGATATATATGAGTTATGTGTATATAGCTCACTGCT        | 1013 |
| Db | 1298 | CCGTGTATTATTAGAACCAAGCGGAGAACAGAGCGGCAACAAAGCCGTAGACTGCTGAT     | 1357 |
| Oy | 1014 | GAATTATGAAAACCTGTGAAGTATTTTATAATGAATATGAAAGACAGATATGATGG        | 1073 |
| Db | 1358 | CAACTACGAGACGCGTGAAGTACTTCACACAGAAAGTACAGGGGGGTGTCTACACGA       | 1417 |
| Oy | 1074 | ATTTTGAAGCGTATGAGACTGTCTTATTGAAAAGTACTTACTCTGGCTATGTGA          | 1133 |
| Db | 1418 | GTGTGTGAAGATGATGAGCGCGGACAGCTTGAAGACTACCTCACTGTGCTGTCA          | 1477 |
| Oy | 1134 | CTTTGTGTCAAGTGTCTATTTTCAAGTGTGCGTTTAAAGCTATATATGTGTGTCCGACGA    | 1193 |
| Db | 1478 | CTTTGGCCGAGAACCCATTTTTCAGACAGCGCATGATCTGATTTATGTATGTGGCGCCAA    | 1537 |
| Oy | 1194 | GGGAATTTGTGCAGGTACCCCTACTGTGTGAGATCTAGTAATGTGGAATGACCTGCTTT     | 1253 |
| Db | 1538 | GGAGATTGTCCCAAGGCAACATGAGGTGTGGAATTTGTGATGTCTAACGCCCTGTCTT      | 1597 |
| Oy | 1254 | TCAGCTTTCATTACCCCTGGAACCTTTCGGAACTGTATATAGAGACTAGACAAGACT       | 1313 |
| Db | 1598 | CGAGCTTCGATCCCGCTCGGTTTTCGGCAGTGTATTATCGTAGGTGCGACAGGCTTT       | 1657 |
| Oy | 1314 | CATGATATGAAACCTTGTTTACTTACTCAAGGTAGACACCCCAATTTAAAGCAAAAGT      | 1373 |
| Db | 1658 | GCTGACATGTGGGCGCATGTGTTCACTGTGATGAACGTGGAACATGACATTTCAACAGCGCG  | 1717 |
| Oy | 1374 | GATGCAATCTCCCTTCAGATCAACACAGACAGCTACCGTGGGCTTGTAAATATGTCA       | 1433 |
| Db | 1718 | CAATCCCAAGCCCTGTTTCTGTGACACACTATCTATCAATTGAGTTCCGACAGTATAG      | 1777 |
| Oy | 1434 | TTTTGATATCAATTGAGGCGCAAGAAAGTCTTATGTGAATATCTTTGAAATCCCTCAGG     | 1493 |
| Db | 1778 | CTTGAAGTACGAACCTGGCAAGCCCAATTTCCGGAGCTCAATTTCAACATATCCCGCGG     | 1837 |
| Oy | 1494 | AAAGAAATGTGCCATTTGTAGAGGTATGTGGGTCAAGGAAAGAACATATATGTAGGCTATT   | 1553 |
| Db | 1838 | TAAAAACGTCCCATTTGTGGCGGCTCTGTCTCAGGAAAAATCGTGCATGTGTGCGCTTCT    | 1897 |
| Oy | 1554 | ATTTGGCTTATATAGGCTCAAAAAGGTGTACATTTATCTTGCTGTCAAAATATATACA      | 1613 |
| Db | 1898 | CTTCCGCTTTCTTTAGCCAACTCTGTATTAAGTGTGATGTGTGGCCAGATATATAGCGC     | 1957 |
| Oy | 1614 | TGTGAGCTTGAAAAGCCTTGTGAGAGGCAGTGGAGTGTACTCAGAGTGTCTCTTT         | 1673 |
| Db | 1958 | CGTGTATTTGGAAGAGCGTGGCAAGGTATATGCAAGTGTGTGCGGCAAGATTTGGTGTATT   | 2017 |
| Oy | 1674 | CCATATATCTATTATTTACAACTCTTATATGAAAAATCATAGTCTTCACTGAGGAAGT      | 1733 |
| Db | 2018 | CCACAAACCATATGAGCAATATATCCACTATGTGCAACTGTCCAAAGTCCCAAGCGGAAGT   | 2077 |
| Oy | 1734 | GTATGCAATGGCAAAATATAGTGTGACTTCATGATGCAATTTCTTGAAATGCCCATGTATA   | 1793 |
| Db | 2078 | GGAGATCTCTCCGCAATGCGTGAATTTGCAATGACTCCATCATATGAGTGTGCGAGCACTA   | 2137 |
| Oy | 1794 | TGACACCCCAATAGGGGAAACAGAGACTCAAGCTTTTCAGAGAGGAAAAAGCAAGTATAGC   | 1853 |
| Db | 2138 | CTCCACTCAGGTATGAGCCAACTGTGATTTAAAGTTATCCGTGTGGGAGAAACAGCGGTTGCG | 2197 |
| Oy | 1854 | ATTTCGAAGAGCCATTTTGAAGACCCCGCAAGTCAATCTTATGATGAAGTACTTTCATC     | 1913 |

| Db                    | 2198   | CATGCTTAGGGCCACTACTGGAAGAACACCCCATCTCTGATCTTTATACAGACGCTTAACAGAC | 2257  |              |                 |
|-----------------------|--|--|---|--------------|-----------------|
| Qy                    | 1914   | GTTAGATTGCATTA CTGAAGAGACTATTCTTGGTCCATGAGAGATGTGTCAAACACAG      | 1973  |              |                 |
| Db                    | 2258   | TCTGGACTCGATTCACCGGACGCTTAACATTCTGACAGCTCTGACCCCGCTTACCTGACAGCG  | 2317  |              |                 |
| Qy                    | 1974   | AACTCTATTTTTCATGACACACATTTGTCAACAGTGGTTGATGACAGTGAATCATTTGT      | 2033  |              |                 |
| Db                    | 2318   | CAC TAGCATTTTGACATGCGCCACCGCTCTCTGACAGTCAAGATCGGACGACGATTTCTGT   | 2377  |              |                 |
| Qy                    | 2034   | CTTGATCAGAGGTAAAGGTAGCGCAACGTCGTACCACCATGTTTGGCT                 | 2081  |              |                 |
| Db                    | 2378   | GCTCGAAGAACGACGCTGTAGGTGAACGTGGACCCATTCGAGAGCTGCT                | 2425  |              |                 |
| RESULT 10             | AF287699   | 2187 bp  | RNA   | linear       | PLN 02-SEP-2000 |
| LOCUS                 | AF287699   |  | Arabidopsis thaliana half-molecule ABC transporter ATM3 mRNA, |              |                 |
| DEFINITION            | complete cds.  |  |   |              |                 |
| ACCESSION             | AF287699   |  |   |              |                 |
| VERSION               | AF287699.1   | GI:9964120   |   |              |                 |
| KEYWORDS              |  |  |   |              |                 |
| SOURCE                | Arabidopsis thaliana.  |  |   |              |                 |
| ORGANISM              | Arabidopsis thaliana.  |  |   |              |                 |
| REFERENCE             | 1 (bases 1 to 2187)  |  |   |              |                 |
| AUTHORS               | Sanchez-Bernandez, R., Mari, S., Dancis, A. and Rea, P. A.   |  |   |              |                 |
| TITLE                 | Functional half-molecule ABC transporters from Arabidopsis: the ATM family   |  |   |              |                 |
| JOURNAL               | 2 (bases 1 to 2187)  |  |   |              |                 |
| AUTHORS               | Sanchez-Bernandez, R., Mari, S., Dancis, A. and Rea, P. A.   |  |   |              |                 |
| TITLE                 | Direct Submission  |  |   |              |                 |
| JOURNAL               | Submitted (17-JUL-2000) Plant Science Institute, Department of Biology, University of Pennsylvania, 3800 Hamilton Walk, Philadelphia, PA 19104-6018, USA |  |   |              |                 |
| FEATURES              | Location/Qualifiers  |  |   |              |                 |
| source                | 1..2187  |  |   |              |                 |
|                       | /organism="Arabidopsis thaliana"   |  |   |              |                 |
|                       | /db_xref="taxon:3702"  |  |   |              |                 |
|                       | /note="ecotype: Col-0"   |  |   |              |                 |
|                       | 1..2187  |  |   |              |                 |
|                       | /codon_start=1   |  |   |              |                 |
|                       | /product="half-molecule ABC transporter ATM3"  |  |   |              |                 |
|                       | /protein_id="AAC09829.1"   |  |   |              |                 |
|                       | /db_xref="GI:9964121"  |  |   |              |                 |
|                       | /translation="MSKGRFRVAPAGLLLRVNIQPPKIPSPSYLSRSDYRLNGES  |  |   |              |                 |
|                       | /transl_except="KTRKIKTSSDSADAMDMKILRTIACIYIMRDPFRFRVIALGFLVAKVAVOV  |  |   |              |                 |
|                       | PTFLKVALDMLASATGASLTTPNATNPLIVFAIPAAVLIGYGLARCSAFNEI   |  |   |              |                 |
|                       | RVNVPTTLEISWISGLAIKFGAAPMTISLSGSIIVTIAVORTYRKAMK   |  |   |              |                 |
|                       | DNDSTRAIDSLINIVETVKYFNNGYAEKXYDCLKYEADALOTQSLAFNFGQST  |  |   |              |                 |
|                       | IFSTALSTAWYLCSGIGINQMVTGDIWMNGLFLQSLPLNFGSVYRETIQSLVDM   |  |   |              |                 |
|                       | KSMQLLEKSDITNTSDAKPLVLCGNIIFENYHFSYLERKILDISFVVPAGKSY  |  |   |              |                 |
|                       | ALVETSGSKSTILRMFRFPDPTGSGNIRIDGDIKEVRIDSLSSIGVVDPTVL   |  |   |              |                 |
|                       | DTIHNHLYGRSLATREEVYEAARAAIHETISNPDKYSYIVERGKILSGGKOR   |  |   |              |                 |
|                       | ILAVTEKSPRLICDRTASLSDSTAEALINLAKLASNRTSFFIAHRLTTMOCDE  |  |   |              |                 |
|                       | ALATFENGVQGHDELLGKSGVADLMTQNSVMDLAATLE"  |  |   |              |                 |
| BASE COUNT            | 598 a  | 479 c  | 483 g   | 627 t        |                 |
| ORIGIN                |  |  |   |              |                 |
| Query Match           | 23.2%  | Score 543.6;   | DB 8;   | Length 2187; |                 |
| Best Local Similarity | 57.5%;   | Pred. No. 1.5e-122;  |   |              |                 |
| Matches 1069;         | Conservative   | 0;   | Mismatches 764;   | Indels 27;   | Gaps 4;         |
| Qy                    | 307  | GGAGTCCACACAGACCCAAAGAGCGTTAAAGATGTTGATCTCGGAAATCATTA            | 366   |              |                 |
| Db                    | 316  | GAAATCAAGACACCTCGCGATTCCAAGCTCAGCTTAGCTGACATGAAGATTTCCGT         | 375   |              |                 |

QY 367 GCAATGCTTTCTTATGTGTGCCCCAAAGACAGCCAGATCTACGAGCTAGAGTTGCCATT 426  
 Db 376 ACGCTTGTGCTATTTGTGATGAGAGACAATCCAGAAATTCGTTTCAGGGGTCAATCCCT 435  
 QY 427 TCGGGGATTTTTGGGTGGTCAAGGCCATGAAATATTGGTGTCCCTTCATGTTTAA 486  
 Db 436 GCTTTGGGTTTCTGTGGGTGCTAAGGCTTGAAATGTCAGATCCCTTCTCTTCAA 495  
 QY 487 TATGCTGTAGACAGCCT-----CAACGATGTCCGGGAAACATGTGTAACCTGAGTG 538  
 Db 496 TTAGCTGTGATTTGGCTTGTCTACCTACCGTACTGAGAGCTTCACTTACCACTTCGCT 555  
 QY 539 ATG-----CACCAATATACAGTTGCAACCATGAGCAACAGAGTTCTGATGGCTAT 588  
 Db 556 GCTACAAACCAACTCTTCTTACTGTTTGTGTACTCTCGCTGCTGTTCTCATGATGATAC 615  
 QY 589 GGTGTATAGAAGCTGAGAGCTGCTTTTAAACGAAAGTTCCAAATGCGATATTGGCAG 648  
 Db 616 GGGATTTGCTCGGACGGGTTCTTGTCTTTTAAATGAACTTCTGACTGCTGTGTTCTTAA 675  
 QY 649 GTAGCCCAAGAAATTCATCCGAAGAAATAGCCAAAAATGCTTTCCTCATCTTCAACAACCTG 708  
 Db 676 GTAGCTCTCCGAACAATACGATCTGTTCTAGAAAGGTGTTCTCAATTTACATGATCTT 735  
 QY 709 GATCTGGGTTTTCACCTGAGAGACAGACGGAGCTTTATCTAAAGCTATTGACAGAGA 768  
 Db 736 GACCTTCGTTACACCTAATGCGAGAAACAGCGGCTTGAAACAGATATTATGATGCTGT 795  
 QY 769 ACAAGGGATATCAGTTTGTCTGATGCTGTTGATTTAAATCTTCTTCCATCATGTTT 828  
 Db 796 AGCCGTGCAATTAATTTATCTCTCACTATGTTTCAATGTTGTTCTTACCAATATG 855  
 QY 829 GAATGATGCTTGTCAAGTGTGTTTGTATTACAAATGCGGTCACAGTTTGCTTGGTA 888  
 Db 856 GAGATATCTATGATATCAGGATATCTGCAATACAGTTTGAAGCTGTTTGCCTGAT 915  
 QY 889 ACCCTGGAACACTTGTGTACATACACAGATTTCAAGTTGAGTCAACGGTGGAGAACT 948  
 Db 916 ACTTCTCTCTCGTTGGATCATATATGCTTTTCAATAGCTGTGACACAGTGGCAACC 975  
 QY 949 AGATTTGAATAGAAATGAACAAAGCAGATATATGATGAGTAATGCTGCTATATGACTCA 1008  
 Db 976 AAGTTTGAAGGCTATAGAACAAAGCTGATATGATGCTACACAAAGGCCATGATGTTCT 1035  
 QY 1009 CTGCTGAATTTAGAACTGTGAAGTATTTAATATGAAGAAATATGAGCAACAGATAT 1068  
 Db 1036 CTGATTAATTTATGAGACTGTCAAAATATTTTAAATGAGTTATGAGGCTGAAAGAT 1095  
 QY 1069 GATGATTTTTGAAGACGTATGAGACTGCTTCATTGAAAAGTACCTACTCTGGCTATG 1128  
 Db 1096 GACCAAGTTCCTGAAAATAATGAGATGCTGCTGCAAACTCAAAAGAGCTTGCTTTC 1155  
 QY 1129 CTGAACCTTTGGTCAAGTGTCTATTTTCACTGTCCGTTTAAACAGCTATATATGCTGCC 1188  
 Db 1156 TTAAACTTTTGAACAAAGATATCATATTTAGTACAGCTTTATCAACAGCAATGTGTGTC 1215  
 QY 1189 AGTCAGGAAATGTGGCAGGATACCTTACTGTGAGATCTAGTAATGTGCAATGAGCTG 1248  
 Db 1216 TCTCAGGGTATCATGAATGGCAGATGACATCGGTATTTGGTTATGTGGAATGAGCTT 1275  
 QY 1249 CTTTTCAGCTTTCAATACCCCTGAACCTTCTGGAAGCTGTATATAGAGAGACTAGACAA 1308  
 Db 1276 CTCTTCAGGCTGTCTCTTCTCTAAATTTTGGGTAGTGTGTTACGTGAAACAAATTCAG 1335  
 QY 1309 GCACCTCATATATGAACACCTGTTTACTCTACTCAAGCTAGACACCCAAATTAAGAC 1368  
 Db 1336 AGCCTGTGTGACATGAATCAATGTTCAGTTGTCTAGAGGAAATCTGACATCACAAT 1395  
 QY 1369 AAGATGATGATCTCCCTTCAGATCAACACAGACAGTACCGGTGCTTGTGAATAT 1428  
 Db 1396 ACAAGTACCGCAAGCTCTT-----GTACTGAAGGTGGAACATTTGTTGAAAT 1449

QY 1429 GTGCATTTTGAAATACATTGAGGGCCAGAAAGTCCCTTAGTGAATATCTTTGAAGTCCCT 1488  
 Db 1450 GTACACTTCAGATTACCTTCCAGAGAGAAAATATTAGATGGAATTTCTTGTGTAC 1509  
 QY 1489 GCAGAAAAGAAAGTGGCCATTGTAGAGTGTGTGGTTCAGGAAAAGCAATAGTAGG 1548  
 Db 1510 GCAGAAAAGAAAGTGGCAATGTGTGTGTACTAGCGGAGCGGCAAGTCAAAATCTTAGG 1569  
 QY 1549 CTATTTATTCGCTTCTATGAGCCTCAAAAGGGTGTGCAATTTATCTTGGTGTCAAAATATA 1608  
 Db 1570 ATGCTATTCAGATTCTTCGACACAGATTTCCGAAATATTAAGATTGACGGCAAGATATA 1629  
 QY 1609 CAAGATGTAGCCTTGAAAGCCTTCGAGGGCAGTGGAGGTGTACTCAGATGCTGTC 1668  
 Db 1630 AAGGAGTGAGACTAGATAGCTCCGAAAGCTCCATAGAGTGTGTCCGCAAGATATCTGT 1689  
 QY 1669 CTCTTCATATATCTATTTATTTATACACCTCTTATATGAAAACATGATGCTTCACTGAG 1728  
 Db 1690 CTCTCAATATGACAAATATTTTCAACATATATATGAGAGCTCTCTCAGCTACAGAAAGAA 1749  
 QY 1729 GAATGTATGAGTGGCAAAATTAAGCTGACCTTGATGATGCAATTTCCGATGCCCAT 1788  
 Db 1750 GAGGTGTATGAGCTGCTCGGCGTGTGCAATTCATGAAACATGATTAATTTCCGTGAT 1809  
 QY 1789 GGAATATGACACCCCAATGAGGGAACGAGACTCAAGCTTTTCAGAGAGAGAAACCAAGA 1848  
 Db 1810 AAATATTCACAGATTCGTTGGGAAAGGGGCTCAAGTTGATGTGTGAGAGAAACAAAGA 1869  
 QY 1849 GTAGCAATTCGAAAGAGCATTTTGAAGAGACCCCACTCATCTATATGATGAAGTACT 1908  
 Db 1870 GTGGCACTTCTCGGACATTTTGAATCCCTGCGATTCGTTGTGTGATGAGGCAACG 1929  
 QY 1909 TCATGCTTATGATGATTTACTGAAAGACATATCTTGTGTCCATGAAAGATGTGTCAA 1968  
 Db 1930 AGTGGCTTGACAGACACAGACAGAGAGATCTGAACCACTCAAGGCACTAGCGAGT 1989  
 QY 1969 CACGAACTTCTATTTTTCATTGTGACACAGATTGTCAACAGTGTGTGATGAGATGAATC 2028  
 Db 1990 AACGAGCATCTATCTTCAATTTGCCCAAGGCTGACCATCGATGCAATGTAGAGATC 2049  
 QY 2029 ATTGTCTTGATCAGGTTAAGTTAGCCGAACGTGTATCCACCAATGTTGCTTGTCTAAC 2088  
 Db 2050 GTAGTACTGGAACACGGGAAGGTGTGTGAACAAAGGCCCCACACAGCAGCTATTGGGAAAA 2109  
 QY 2089 CCTCATATGATCTTATTCAGAAATGTGGCATACACAGAGCAGCGGTGCGAAGCAATGAT 2148  
 Db 2110 TC---AGGGAGATATGCGCAGCTATGAGACAAACAAACAGCAGCGGTGATATGCTTGAT 2166

RESULT 11  
 AF360334  
 LOCUS  
 DEFINITION  
 Arabidopsis thaliana putative ABC transporter protein (At5g58270)  
 mRNA, complete cds.  
 ACCESSION  
 AF360334.1 GI:13430843  
 VERSION  
 FLI CDNA.  
 KEYWORDS  
 SOURCE  
 Arabidopsis thaliana.  
 ORGANISM  
 Arabidopsis thaliana.  
 Buxarjocia; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 2373)  
 REFERENCE  
 AUTHORS  
 Yamada, K., Liu, S.-X., Sakano, H., Pham, P.-K., Banh, J., Chung, M.-K.,  
 Goldsmith, A.D., Lee, J.-M., Quach, H.-L., Toriumi, M., Yu, G., Bower, L.,  
 Carinici, P., Chen, H., Cheuk, R., Hayaishi, Y., Ishida, J.,  
 Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,  
 Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T.,  
 Satou, M., Seki, M., Shim, P., Southwick, A., Shinzaki, K.,  
 Davis, R. M., Ecker, J. R. and Theologis, A.  
 TITLE  
 Arabidopsis Full Length cDNA Clones  
 JOURNAL  
 Unpublished  
 REFERENCE  
 2 (bases 1 to 2373)



```

Db 1473 GTACCTTCAGTACCTTCCAGAGAGAAAATATTAGTAGAATTTCTTTGTGTAACCG 1532
Qy 1489 GCAGGAGAGAAAGTGGCCATTGTAGAGGTAGTGGGTGAGGAGAAAGCAATAGTGAGG 1548
Db 1533 GCAGGAGAGAAAGTGGCAATGTTGTACTAGCGGAGACGGCAAGTCAACAATCTTAGG 1592
Qy 1549 CTATTATTCCGCTTCTATAGACCTCAAAAAGGTAGCATTTATCTTGTGTCGTAATAATA 1608
Db 1593 ATGCTATTCAATCTTCTGACACAGATTCGGAATATATAGAGATTAGCGGCAAGATATA 1652
Qy 1609 CAAGATGAGCCTGAGAAAGCCTTCGAGGAGAGTGGAGTGTACCTCAGAGATGCTGC 1668
Db 1653 AAGGAGTGAACCTAGATAGCCTCCGAAGCTCCATAGAGTGTGCGCAAGATATCTGT 1712
Qy 1669 CTCTCCATTAATTAATTAATTAACAACCTTTATATAGAAACATCAGAGCTTCAAGT 1728
Db 1713 CTCTTCATGACACAAATTTTCAACAATATATATATATATATATATATATATATATATAT 1772
Qy 1729 GAAGTGTATGCAAGTGGCAAAATTAAGTGGACTTCATATGCAATTTCTTGAATGCAAT 1788
Db 1773 GAGGTGTATGAGGCTGCTCGGCGTGTCTCAATTCATGAAACATCAGTAATTTCCCTGAT 1832
Qy 1789 GGATATGACACCCAAAGTATGAGGAGAGAGACTCAAGCTTTTCAAGAGAGAGAGAGAGAG 1848
Db 1833 AAATATTCACAGATGTTGGGAGAGAGGAGGCTCAAGTGTAGTGTGAGAGAGAGAGAGAG 1892
Qy 1849 GTACCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1908
Db 1893 GTGGCACTGCTGCGGACATTTTGAATCCCTGCGATTCCTGTGTGTGTGTGTGTGTGTGT 1952
Qy 1909 TCATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1968
Db 1953 AGTGGCTGAGCAGACACAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 2012
Qy 1969 CACGAGACTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2028
Db 2013 AACGAGATCTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 2072
Qy 2029 ATTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2088
Db 2073 GTAGTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2132
Qy 2089 CCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2148
Db 2133 TC--AGGAGATATGCGCAGCTATGACACACAAAAACAGACGGTGTGATGCTTGAT 2189

RESULT 12
AF287697 2037 bp mRNA linear PLN:02-SEP-2000
LOCUS Arabidopsis thaliana half-molecule ABC transporter AT1M1 mRNA,
DEFINITION complete cds.
ACCESSION AF287697
VERSION AF287697.1 GI:9964116
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
AUTHORS Sanchez-Fernandez, R., Mari, S., Dancis, A. and Rea, P. A.
TITLE Functional half-molecule ABC transporters from Arabidopsis: the ATM
JOURNAL family
REFERENCE 2 (bases 1 to 2037)
AUTHORS Sanchez-Fernandez, R., Mari, S., Dancis, A. and Rea, P. A.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2000) Plant Science Institute, Department of
Biology, University of Pennsylvania, 3800 Hamilton Walk,
Philadelphia, PA 19104-6018, USA
FEATURES
Location/Qualifiers

```

```

source 1..2037
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/notes="ecotype: Col-0"
1..2037
/codon_start=1
/product="half-molecule ABC transporter AT1M1"
/protein_id="AG09827.1"
/db_xref="GI:9964117"
/translation="MMRGRFLPLSRALYRLRSDHHHSFSSIKRSSIORSPAINA
FTDPSPSPVNARVFPSTSTSPNOOTKASSKILRITISYLMKNDLRF
VIAALCLIGAKRNQVPELPFLSDILSSYSSTTSPNPLAFAFRPSVLFG
GIARSSSAFNEILRTAVESKSLRITRSRVKLSLHDLRLYHLAREGALMLID
RGSRAINTLSAMVFNVPITLIESWVTGILANFGPVALYDLSVGSYIAFTLVV
QYRTERKAMNDNDASTRAIDSLVNEIVKFNNEDEARLYDLFQSLPFLYG
OSLIAFLDGOSFIFSTALSTSVLSOSIENGEMTVGLVMVNGSLPENVHPSYLP
GYRRETVQGLVPMKSLFOLLERSDIDGMDTEKLPYLVRGSLSPENVHPSYLP
KIIDGISFEVPACKSVATVGSSSGSTLLRMTFRFPDSDGNVRLDGODIKETLES
LNSCIGVFPQDVLVENDTLFHNHYGNLSATEEVEVDARAVIHDTIMKPEPKYSTA
VBERGLMSGGEKORVALARALPKSPAILLCEATNALDSKTEAEIMKTFRLSNRT
CIFIAHRLTAMQDEIIVMEKRVKEKTHQVILBSGRYAKLMTQNSTLEV"
BASE COUNT 576 a 407 c 439 g 615 t
ORIGIN
Query Match 23.2%; Score 543.4; DB 8; Length 2037;
Best Local Similarity 57.2%; Pred. No. 1.7e-122;
Matches 1026; Conservative 0; Mismatches 761; Indels 6; Gaps 2;
Qy 350 CTCGGAATATCATAAAGCAATGCTTCTATGTGTGGCCCAAGACAGCGCATGCTAC 409
Db 239 CTAAAGAGATCTCCGACAGATTTCAAGCTATCTATGATGAAGAACACCCAGAACTCC 298
Qy 410 GAGCTAGATGTCATTTCCGCTGGATTTTGGGTGTGCAAGAGCCATGAATATTTGTG 469
Db 299 GATTGAGATCATGCGCTTTAGCGGTCTCATAGAGCTAAGTTCTTGATGTTCAAG 358
Qy 470 TTCCCTTCATGTTTAATATATGCTGTAGACAGCTCAAC--AGATTCGGAGAAACATGC 526
Db 359 TTCCCTTCCTTCAATTTATTCATGATTTACTCTCTTCATATTTCACTTCAACAATFCA 418
Qy 527 TGAACTGATATGACCAATAATACATTTGCCAACATGGAACAGCGTTCTGATTTGGCT 586
Db 419 CTGATTCMAATCCCTATCTATCTTGTGCTTTCCTCTTCCGTTCTCATTTGGCT 478
Qy 587 ATGATGATCAAGAGCTGAGCTGCTTTTAAACGAAGTTGCAAGATGAGATTTGGCA 646
Db 479 ATGGCATTTGCTCGATCTGATCTTCAAGCTTTTAAAGCTTCGTACTGCTGTTTCTCCA 538
Qy 647 AGGTAGCCAGAAATTCATTCGGAAGATAGCCAAATATGTCCTTCTTCATCTTCAACACC 706
Db 539 AAGTATCTCCGACAAATAGATCTGTTCTGTAAAGTCTCTGCAATTTACATGATC 598
Qy 707 TGGATCTGGGTTTTCACCTAGACACAGCGGAGCTTTATCTAAGGCTATTTGACAGAG 766
Db 599 TTGATCTTGGTACCAATTTGAATCGAAGAACAGAGAGCTTTGAATAGATCATTTGACCGTG 658
Qy 767 GAAAGAGGGGTATCAGTTTGTCTGAGTGTGATTTAAATCTTCCATCTTCAACATGT 826
Db 659 GAAGTGGCGGATTTACACAGATCTCTCTGCTATGTTTCAATGTTGTTTCAACATCT 718
Qy 827 TTGAAGTATGCTTTCAGTGTGATTTTGTATTAACAATGCGGTGCCAGTTGCTTTGG 886
Db 719 TGGAGATATCAATGATGATATTTAGCATATCAATTTTGAACGCTGTTTGGCTTTGA 778
Qy 887 TAACCTTGAACACTTGTGATATACACAGATTTACAGTGTGACACAGCGTGGAGAA 946
Db 779 TCACTTCTCTTTCGTTGATATCATATACATCTTTCACATTTGTTGACACAGATTAAGAA 838
Qy 947 CTGATTTAATATGAATATGAACAAAGACAGATATATGATGAGGAAATGCTGTATAGACT 1006
Db 839 CTAAAGTTTAAAGAGCTATAGACCAAGCTGATATATGATGACACAAAGAGCTATTGATT 898
Qy 1007 CACTGTGATATTAAGAACTGTAAGATTTTAAATATGAAGATATGAAGACAGAGAT 1066

```



```

Db      899  CTCTGTAACATACGAGACTGTTAGTATTTTAAACAAGAGATTACGAAGCGAGAAAGT  958
      1067  ATGATGATTTTTGGAAGAGTATGAGACTGCTTCAATTGAAAGTACTCTTGCGCTA  1126
      959  ACGAGATTTTACTCGGAATATGGAAGATGCGCATTTGCAAACTAGAAAGATCTTGGCT  1018
      1127  TCGTAACCTTGTGCAAGTGTCTATTTTCAGTGTGCGTTTAAACAGCTAATATGTCGTCG  1186
      1019  TTTTGAATTTTGGACAGAGTTTATATTTAGTACAGCTTTTGCACATCAATGAGTGTGT  1078
      1187  CCAATCAGGGAATTTGTGCAAGTACCCTTACCTTTGAGATCTAGTATGTGTAATGAC  1246
      1079  GTTTCAGAGGATTTGATGATGAGAGATGACAGTTGGATTTGGATGTGGAATGTC  1138
      1247  TGCCTTTTCAGCTTCAATTAACCCGAACTTCTGGGAACCTGATATAGAGAGCTAGAC  1306
      1139  TTCTCTTCAATGTCTCTCTCTTCTTCTTCTGCGTGTGTTTACCGTGAACCGTTC  1198
      1307  AAGCACTCATGATATGAAACACCTTGTACTCTACTCAAGGTAGACACCAATTTAAAG  1366
      1199  AGGCTCTGTATATGAAATCATTTGTTCCAGTTGCTAGAGAGAGATCCGATATTGGAG  1258
      1367  ACAAGTGTGSCATCTCCCTTCAGATCAACAACAAGCT--ACCGTGGCCCTTTC  1423
      1259  ATAAAGATACAGAGACAAAGCTACCTCTTCTGTTCTGAGAGTGGAGCATCAGTTTTC  1318
      1424  ATAAATGATTTTGAATATCATTGAGGCGCAGAAAGTCTTATGTAATCTCTTTGAAG  1483
      1319  ACAATGTATATTTAGTTATTCGCAAGAGAAAGTTTGGATGGATTTTCAATTTGAG  1378
      1484  TCCCTGCAAGAAAGAAAGTGGCATTGTGAGAGTGTAGGAGTGTGAGGAAACACAATAG  1543
      1379  TACCGGCAAGGAAAGTGTAGGATTTGTGTAGTATGAGGAAAGCGCAATTCAGCATTC  1438
      1544  TGAGCTATTTATTTGCTTCTATGAGCTCAAAAGGATGACATTTATCTTGTGCTCAA  1603
      1439  TTAGAATATTTTCAAGTCTTTTGAACAAGATTTGTGATATGTAGATTTGATGTCAG  1498
      1604  AATTACAAATGTGAGCTTGGAAAGCTTGGAGGCGAGTGGAGTGGTACTCAGATG  1663
      1499  ATATTAAGAAATGACATAGAGAGCTTTCGAAGTGCATTTGAGTGTGCTCAAGATA  1558
      1664  CTGTCTCTTCCATATCTATTTATTAACAACCTCTTATATGAAACATCAGTGTCTCAC  1723
      1559  CTGTCTCTTCCATGACACAAATTTTCAACAATTCACATTCATATGAAACCTTTGAGCAAG  1618
      1724  CTGAGGAAGTATGCAAGTGGCAAAATTAAGCTGACCTTCATGATGCAATTTCTGAATGC  1783
      1619  AAGAAGAGGTATACATGACAGCTAGCGCAGAGTGAATTCACATACATGAAATTTTC  1678
      1784  CACATGATATATGACACCCAAAGTGGGAAACGAGACTCAAGCTTTCAGAGAGAGAAAGC  1843
      1679  CGGATTAATATTTGCAACCGGCGTGGAGAAAGGGCTGATGTGAGTGTGAGAAAAAC  1738
      1844  AAAAGTGAATGTGCAAGAGCCATTTTGAAGAACCCCCAGTCAATCTATGATGAAG  1903
      1739  AAGAGTGCACCTTGTCTGTCTGCTTTCTGAATTCACACAGCATTTCTTGTGTGACGAAG  1798
      1904  CTACTTCATCTGTTAATTCGATTAATGAGAGACTAATTTCTTGTGCAATGAAGATGTGC  1963
      1799  CAACGAATGCGTTGATATTAACCGAGGACAGATATGAAAGCGTTCAAGTGGCTAG  1858
      1964  TCAAAACAGAACTTCTATTTTTCATGCAACAGATTTTCAACAGTGTGTTGTCAGATG  2023
      1859  CGAGTAAACGAAATGCAATTTTTCATTTGCAATAGGCTGACCACTGATGTCATGTGATG  1918
      2024  AAATATGTTCTTGTGATCGGGTAAAGGTAGCCGAAGTGTGATCCCAACATGTTTGTCTG  2083
      1919  AATATATGTAATGAGAAAGGAAAGTGTGAGAAAAAGGACACACAAAGTGTGTGG  1978
      2084  CTACCTCATATGATCTATTTCAGAATGTGCAATACAGAGACAGCCGTGTG  2136

```

```

Db      1979  AGAAATCTGCGAGATACGCTAATTAATGACACACAAACAAACGATACACTGAG  2031
      RESULT 13
      ATH272202
      LOCUS
      DEFINITION
      Arabidopsis thaliana mRNA for mitochondrial half-ABC transporter
      (STAI gene).
      ACCESSION
      AJ272202.1 GI:9187882
      VERSION
      AJ272202.1 GI:9187882
      KEYWORDS
      mitochondrial half-ABC transporter; STAI gene.
      SOURCE
      ORGANISM
      Arabidopsis thaliana
      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
      Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsis.
      1 (bases 1 to 2452)
      AUTHORS
      Kushnir, S., Babiychuk, E., Storozhenko, S., Davey, M. W.,
      Papenbrock, J., De Rycke, R., Engler, G., Stephan, U. W., Kissal, G.,
      Lill, R. and Van Montagu, M.
      TITLE
      A Mutation of the Mitochondrial ABC Transporter Stai Leads to
      Dwarfism and Chlorosis in the Arabidopsis Mutant stearik
      JOURNAL
      Plant Cell 13 (1), 89-100 (2001)
      REFERENCE
      Kushnir, S.G.
      2 (bases 1 to 2452)
      Direct Submission
      Submitted (14-FEB-2000) Kushnir S.G., Plant Genetics, VIB,
      University of Gent, Ledeganckstraat 35, B9000, BELGIUM
      FEATURES
      source
      1. .2452
      /organism="Arabidopsis thaliana"
      /db_xref="taxon:3702"
      27..2213
      /gene="STAI"
      27..2213
      /gene="STAI"
      /function="Fe/S cluster biogenesis"
      /codon_start=1
      /evidence=experimental
      /product="mitochondrial half-ABC transporter"
      /protein_id="CA97048.1"
      /db_xref="GI:9187883"
      /db_xref="SPTRMBL:09LF78"
      /translation="MSRGSFRVAPGSLPCRVLQPOPKIPSPSYLSRSPYRLHNGFS
      NYTRNSIRTPVINAFLSDNPSPPSPGPRFVORSSMLNRLSTSPNDQDTT
      KTKRKITSSDSADMDKILKTLAGLYLMRNPFRPRVIALGLFVIGAKYLNQV
      PFLFKLAVMDLASATGASLTTPAATNPLTLVATPAVALIGYGIARTGSAPNEL
      RTAVFSKVALRTIRSVSRKVFSLHDLDRYHLSRBTGGLNRIIDGSRXINFLSAM
      VPAVVPFLIELISWGSGLIAYKFGAFAAMITSLVGSYIVETTLAVOTMRTEFSAMNKA
      DNDSTRAIDSLINVEYKYFNNGEYAEKYDQFLKKYEDPAALOTORSIAFLNFGOSI
      IESTALSTANVLCSQGINMGQMTVGDVWVNGSLFOLSLPINFLGSVYRETTOSLYDM
      KSMFQLEEKSDITNTSDAKPLVLKGNIEFEVHPSYLPERKILDGISFVPAQSV
      AIVGTSGSGSTLIRMLFRFEDDSGNIRIDGQDIEVRIDSRSSIGVPODTVLFN
      DTFPHNHYRLSTAEVEEARRAAIIHTISNFPDKYSTIVGEGKLKXSGEKGQV
      ALRTFLKSPAILCDEATSEALSTTEALINLKLALASRTSIFLAHRLTTAMQODE
      IIVLENGKVVQEGDELLKSGRYAQLMTQGNSSVMDMAIKLE"
      BASE COUNT
      703 a 519 c 537 g 693 t
      ORIGIN
      Query Match 23.0%; Score 540.4; DB 8; Length 2452;
      Best Local Similarity 57.4%; Pred. No. 9,5e-122;
      Matches 1067; Conservative 0; Mismatches 766; Indels 27; Gaps 4;
      Db      307  GGACTCCACACAGACCCAAAGAGGTTAAAGATGTTGACTCGGAAAATCATATAA  366
      342  GAAATCAAGACAACTCGTCGATTCAAGCTCAGCTATGCTGACATGAAGATCTCCGT  401
      367  GCAATGCTTTTATATGTGTGCGCCCAAGACGAGCAATCTACAGACTGATGTCATT  426
      402  ACGCTGTGCTATTTTGTGATGAGACATCAAAATTTCTTTTCAGGCGATGCT  461
      427  TCGCTGGATTTTGGGTGTGCAAGGCCATGATATTTGTGTTCCCTTCATGTTTAA  486

```



```

Db      462 GCTTGGGTTTCTCGTGGGTGCTAAGGTCATGATGCCAAGTCCCTTCTCTTCAA 521
Qy      487 TATGCTGAGACAGCT-----CAACGAGTGGGGAACAATGCTGAACCTGAGTG 538
Db      522 TTAGCTGTGATGATGCTTGTCTTCACTACCGTACTGAGCGCTCACTTACCACTTGTGCT 581
Qy      539 ATG-----CACCAATATACAGTTGCAACCATGCGCAAGCAGTCTGATGGCTAT 588
Db      582 GCTACAAACCCCACTCTTCTTACTGTTTTTGTGACTCTGCGTGCTTCTATTGGATAC 641
Qy      589 GGTGATATAAGAGCTGAGAGCTGTTTTTTTAAACGAAGTTCGAATGCAATATTTGGCAAG 648
Db      642 GGGATGCTGGACGGGTTCTTGTCTTTTATGAACTTCTACTGCTGTCTTCTAAG 701
Qy      649 GTACCCCAAGATTCGAATCCGAAGAATAGCCAAATGCTTTCTCATCTTCAACAACCTG 708
Db      702 GTAGCTCTCCGAACAAATACGATCTGTTCTTGAAGAGTGTCTCACTTATCAATGATCTT 761
Qy      709 GATCTGGGTTTCACTGAGCAGACAGACGGAGCTTATCTAAGGCTATTGACAGAGA 768
Db      762 GACCTTGCTTACCACTAATGCGAAGAACAGCGGCTTGAACAGGATTTATGATCGTGT 821
Qy      769 ACAAGGGGATCACTTTTGTCTGAGTCTTTGGTATTATCTTCTTCCATCATGTTT 828
Db      822 AGCGTGCAATTAATTTATCTCTCACTATGTTTCAATGTTTCTTACCATATATG 881
Qy      829 GAAGTATGCTGTCAAGTGTGTTTTGTATTAACAATGGGTGCCAGTTGGCTTGTGTA 888
Db      882 GAGATATATATGATATCAGTATACGTGACATACAGTTGAGCGCTTTTGGCTGAGTT 941
Qy      889 ACCCTTGGAACACTGTGATACATACACAGATTCACAGTTGAGTACACAGGTGAGAACT 948
Db      942 ACTTCTCTCTCGTTGATATATATATGTTTCACTTACTGATGACACATGTGGGAAC 1001
Qy      949 AGATTATGATATAAATGACAAAGCAGATATATGATGAGTAAATGCTGTATAGACTCA 1008
Db      1002 AAGTTATGATAGGCTATGACAAAGCTGATATGATGATACACAAGGCGCATGATCTT 1061
Qy      1009 CTGCGAATTTGAAACGTGAAAGTATTTTATATATGAATGAATGAAAGCACAAGATAT 1068
Db      1062 CTGATTAATTAATGAGACTGTGAAATATTTTAAATGAAAGGTTATGAGCGTGAAGAT 1121
Qy      1069 GATGATTTTGAAGAGCGTATGAGACTGCTTCAATTGAATAATACCTTACTCTGCTATG 1128
Db      1122 GACAGATTCCTGAAATAATATGAAAGATGCTGCTGCTGCAATCAAAAGATCTTGTTC 1181
Qy      1129 CTGAATCTTGTGCAAGTGTATTTTCAAGTGTGCTTAAACGCTATATATGATGCTGCC 1188
Db      1182 TTAAACTTTGACAAAGATATATTTATGATCACTTTATCAACAGCAATGTGTGTGC 1241
Qy      1189 AGTAGGAATGTGGACGAGTACCTTACTGTGAGATCTAGTAATGAAATGAGACTG 1248
Db      1242 TCTGAGGATATCATGATATGCGACATGACATGCGGTGATTTGGTTATGTTAAAGACTT 1301
Qy      1249 CTTTTCAGCTTCAATTAACCTGAACTTTCTGGAACTGTATATAGAGACTAGACAA 1308
Db      1302 CTCCTTCACTGCTCTCTCTTAATTTTGGTAGTGTTAACCGGAACAATTCAG 1361
Qy      1309 GCATCATATGATATGAAACCTTGTTACTTACTCAAGGTAGACACCCAAATTAAGAC 1368
Db      1362 AGCCTTGAGCATGAATATCATGTTCAGTTGCTAGAGGAGAATCTGACATACAAAT 1421
Qy      1369 AAATGATGAGCATCTCCCTTCAGATCACACACAGACAGTACCGTGGCTTGTGATAT 1428
Db      1422 ACAAGTACCGAAAGCTCTT-----GTACTGAAGGTGAAACATTGAGTTTGAAT 1475
Qy      1429 GTGCAATTTGATATCATTTGAGAGGCGCAAGAACTCTTAGTGAATATCTTTGAAGTCCCT 1488
Db      1476 GTACACTTCAAGTTACTTCCGAGAGAAAATATATAGATGAATTTCTTTGTGTACCG 1555
Qy      1489 GCAGAAAAGAAAGTGGCAATTTGAGAGGTATGAGGTGACGGGAAACACAAATAGTAGG 1548

```

```

Db      1536 GCAGAAAAAGTGTGCAATCTTTGTACTAGCGGAGCGGCAATTCATTTAGG 1595
Qy      1549 CTATATTTGCTCTATGAGCCCTCAAAAGGATGATTTATCTTGGTCAAAATATA 1608
Db      1596 ATGCTATTCAGATTTCTTGCACAGATTTCCGAAATATTAAGATTGACGGCAAGATATA 1655
Qy      1609 CAAGATGAGCCCTGAAAGCTTCCGAGGGCAGTGGAGTGTACTCAGATGCTGTC 1668
Db      1656 AAGGAATGAGACTAGATAGCTCCGAAGCTCCATAGAGTGTGCGCAAGATATCTGT 1715
Qy      1669 CTCTTCATTAATACTATTTATTAACCTCTTATATGAAACAATCACTGCTTACCTGAG 1728
Db      1716 CTCTTCATGACACAAATATTTCAACACATCATTTATGAGCTCTTCACTACGAAGAA 1775
Qy      1729 GAAGTATGACAGTGGCAAAATTAAGTGTGACTGATCATGATGCAATTTCTTGAATGCCAT 1788
Db      1776 GAGGTATATAGGCTGCTCGGCGTGTGCAATTCATTAACATCAATATTTCCCTGAT 1835
Qy      1789 GGATATGACACCCAAAGTGGGAAACGAGACTCAAGCTTTCAAGAGAGAAAAGCAAGA 1848
Db      1836 AAATATTCACAGATCGTTGGGAAAGGGGCTCAAGTTGATGAGTGAAGAAACAAAGA 1895
Qy      1849 GTAGCATTTGCAAGAGCCATTTTGAAGAACCCCACTCATCTTATGATGAAGCTACT 1908
Db      1896 GTGCACTTGTCTGGACATTTTGAATCCCTGCGATTTGTTGTGATGAGGCAACG 1955
Qy      1909 TCATCTTATGATGATTAATGAAAGACTATTTCTTGTGTCATGAGAGATGTTGTCATA 1968
Db      1956 AGTGGCTGACACAGACAGAGGACAGATACTGAACGACTCAAGGCACTAGCGAGT 2015
Qy      1969 CACGAACCTTCTATTTTCACTTGACACAGATTTGCAACAGTGTGTTATGACATGAATAT 2028
Db      2016 AACGAGCATCTATCTTATTTGCCACAGGCTGACCATCGATGCAATGATGAGAGATC 2075
Qy      2029 ATTGCTTGATGAGGTATGATGACCGGAACGTGTATCCCAATGCTTGTCTTCTTAC 2088
Db      2076 GTAGTACTGGAACCGGAGAGGTGTGGAACAAAGCCCCACGACGCTATTTGGGAAA 2135
Qy      2089 CCTCATATGATCTATTCAGAAATGTGSCATACACAGACGAGCGGTGCGAAGCATGAT 2148
Db      2136 TC--AGGGAGATATGCGCAGCTATGAGACAAACAAACAGCAGCTGATATGCTTGAT 2192

```

```

RESULT 14
AF287698      2043 bp      mRNA      linear      PLN 02-SEP-2000
LOCUS       AF287698      Arabidopsis thaliana half-molecule ABC transporter ATM2 mRNA,
DEFINITION complete cds.
ACCESSION   AF287698
VERSION     AF287698.1 GI:9964118
KEYWORDS
SOURCE      Arabidopsis thaliana.
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 2043)
AUTHORS     Sanchez-Fernandez,R., Mari,S., Dancis,A. and Rea,P.A.
TITLE       Functional half-molecule ABC transporters from Arabidopsis: the ATM
            family
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 2043)
AUTHORS     Sanchez-Fernandez,R., Mari,S., Dancis,A. and Rea,P.A.
TITLE       Direct Submission
SUBMITTER   Submitted (17-JUL-2000) Plant Science Institute, Department of
            Biology, University of Pennsylvania, 3800 Hamilton Walk,
            Philadelphia, PA 19104-6018, USA
FEATURES
            source
            1..2043
            /organism="Arabidopsis thaliana"
            /db_xref="taxon:3702"
            /note="ecotype: Col-0"
CDS
            1..2043

```



VERSION AC018231.1 GI:6552960  
 KEYWORDS HTG; HTGS\_PHASE2.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 50701)  
 AUTHORS Adams,M. and Venter,J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 COMMENT This sequence was identified as CDM:10213948 by the submitter.  
 For more information on this record e-mail to fly@celera.com.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 FEATURES  
 source  
 1. .50701  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 BASE COUNT 13744 a 12081 c 12135 g 12741 t  
 ORIGIN  
 Query Match 20.5%; Score 480.6; DB 2; Length 50701;  
 Best Local Similarity 59.6%; Pred. No. 5.2e-107;  
 Matches 810; Conservative 0; Mismatches 549; Indels 0; Gaps 0;  
 QY 587 ATGCTGATCAAGAGCTGGAGCTGTTTAAAGAAAGTTGAAATGCAATGCAATTTGGCA 646  
 DB 11810 ATGGCATGCTGAGCCGAGCTGGCGGGCTTTAATGAGTTCCGGAATGCAATGCTTGGCA 11751  
 QY 647 AGTAGCCCAAGATTCATCCGAAAGATAGCCAAATATGCTTCTCATCTTCAACACC 706  
 DB 11750 AGGTGGCCCACTCTGATCCGAAAGATCGCCAGCAATGTTCTTGATCTGCAACACC 11691  
 QY 707 TGGATCTGGGTTTCACTGAGCAGACAGACGAGAGCTTTATCTAAGCTAATGACAGAG 766  
 DB 11690 TGGATCTGGGCTTCCACTGAAACAAACAACTGAGAGCTGTCAAAAGCACTGATGAG 11631  
 QY 767 GAACAAAGGGATCAAGTTTGTCTGAGTGTCTTGGTATTAATCTTCTTCCATCATGT 826  
 DB 11630 GATGAGAGGGGATTAACCTTGTGCTCCGCGATGCTTCAACATTTGCTCCCATCTCT 11571  
 QY 827 TTGAAGTATGCTGTGAGTGTGTTTGTATTAACAATGGGGGCCAAGTTGCTTTGG 886  
 DB 11570 TTGAGCTGGCCCTGCTGATCCAGTATCCGAGGAGTGAAGTGGCCCTGCTGCTGGTG 11511  
 QY 887 TAACTCTTGAACACTTGTATCATATACAGACATTCACAGTTGACAGTCAACGAGTGAAG 946  
 DB 11510 TCACGATGGGCTGGCTGGCTGATATACGCTGCTACACTGAGTGAACCAAGTGGGCA 11451  
 QY 947 CTAGATTTAGAAATAGAAAGCAAGCAATTAATGATGAGTAAATGCTGTATAGACT 1006  
 DB 11450 CCCAGTTCCGATTTATTAAGAACAGGGGAGAAAGAGCCGAGCAACAGGCCGTAGACT 11391  
 QY 1007 CACTGCTGAATTAAGAACTGAAATTTTAATTAATGAAGATAGAAAGACAGAGAT 1066  
 DB 11390 CGCTGATCAACTACAGAGCGGTGAAGTACTTCAACAAAGAAAGTACGAGGCGGGTGGT 11331  
 QY 1067 ATGATGATTTTGAAGACGATAGACTGCTTCAATGAAGAAAGTACCTTCACTGAGCTA 1126  
 DB 11330 ACAACGAGGTGCTGAAGAAAGATAGAGGGGCGACACTGGAAGACTAGCTCCAGTCCGCTC 11271  
 QY 1127 TGTGAACCTTTGCTCAAGTGTATTTTCACTGCTGGTTTAAAGCTAATAATGCTGCTCG 1186  
 DB 11270 TGTCAACTTTGGCCAGAACGCCATTTTTCAGCAGCGCACTGATCTGATTAATGTAATGG 11211  
 QY 1187 CCACTCAGGGAATTTGGCAGCTACCTTACTGTTGAGATCTAGTAATGTTGAATGAC 1246  
 DB 11210 CCGCAAGAGATTTGCCAAGGCAACATGACGTTGGAGATTTGGTATGATCAACGCC 11151

QY 1247 TGTCTTTGACCTTTCATTAACCCCTGAACCTTTCTGGGAAGTATATAGAGACTAGAC 1306  
 DB 11150 TGTCTTTGACCTTTCATTAACCCCTGAACCTTTCTGGGAAGTATATAGAGACTAGAC 11091  
 QY 1307 AAGCACTCATAGATAGAAACACCTGTTTACTACTCAAGTGAAGACCAAAATTAAG 1366  
 DB 11090 AAGCTTTGCTGACATCGGGGCCATGTTTCACTGTTGAAGAGTGAAGATGACATTTGCA 11031  
 QY 1367 ACAAGTATGATGATCTCCCTTCAGATCAACCCAGACAGACGATACCGTGGCTTTGATA 1426  
 DB 11030 CAGCCGCCAATGCCCAGACCCCTGTTGAGACACCACTAATCATATGATGATTCGCCA 10971  
 QY 1427 ATGTGCAATTTTGAATATGATGAGGCGCAAAAGTCTTATGTAATTCCTTGAAGTCC 1486  
 DB 10970 ACGTGAACCTTGAATGAACTGGCAAGCCCAATTTCCGGGACCTCAGATTTCAACATAC 10911  
 QY 1487 CTGAGGAAGAAAGAGGCGCATGTTGAGAGTGTGCTGAGGAAAGCAATATGTA 1546  
 DB 10910 CCGCGGTAAACGTCGCAATGTTGGGGGCTGTGCTCAGGAAATTCGTGATGTTGC 10851  
 QY 1547 GGTATTAATTTGCTCTATGAGCTCAAAAGGATAGCAATTTATCTTGTGTCATAATA 1606  
 DB 10850 GCTTCTCTCCGCTTCTTTGAGCCAACTCTGTTAAGTGTGATGGTGGCCAGATA 10791  
 QY 1607 TACAAGATGATGAGCTGGAAGAGCTTTCGAGGCGAGTGGAGTGTACTTCAGAGTCTG 1666  
 DB 10790 TCAGCGCGGTGATTTGAGAGCCTGCGCAAGGTATTTGACATGATGTCGCCAGATTCGG 10731  
 QY 1667 TCTCTTCCATTAATATTTATTAACCTCTTATATGAAACATAGTGTTCACCTG 1726  
 DB 10730 TGTATTTCCAAACACCATGAGACCAATATTCATATGGAACCTGTCCAGTCCACG 10671  
 QY 1727 AGGAAGTATGCAAGTGGCAAAATAGCTGACTGATGATGCAATCTTGAATGCCAC 1786  
 DB 10670 CAAAGTGCAGATATGCTGCCCGCATGCTGATTTGATGATCTCATATAGCTGGCCAG 10611  
 QY 1787 ATGATATGACACCCCAATGAGGGAACGAGACTCAAGCTTTGAGAGGAAAGCAAA 1846  
 DB 10610 GACAGTACTCAGTCAAGTATGAGGCAAGTGTGATTAAGTTATCCGGTGGCGAAGGAGC 10551  
 QY 1847 GAGTAGCAATTTGCAAGAGCCATTTTGAAGACCCCGGACATATCTTATGATGAACCTA 1906  
 DB 10550 GCGTTGCAATCGCTAGGGCCATCTGAAAGAACACCCCATTTGATCTTGAACGAGCTA 10491  
 QY 1907 CTTCATGTTAGATGATGATCTGAAAGACTATTTCTTG 1945  
 DB 10490 CAAGCACTGAGCTGATCAACGAGCATGTGAGTGTG 10452

Search completed: February 15, 2003, 03:21:32  
 Job time : 6092 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2003, 00:34:00 ; Search time 480 Seconds  
(without alignments)  
11001.950 Million cell updates/sec

Title: AF133659  
Perfect score: 2345  
Sequence: 1 ATGGCGCTGCTCGCATGCA.....TATTAAAAATCATCACTT 2345

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :  
1: /SID2/gcgdata/geneseq/geneeqn-emb1/NA1980.DAT:\*  
2: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT:\*  
3: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT:\*  
4: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT:\*  
5: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT:\*  
6: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT:\*  
7: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT:\*  
8: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT:\*  
9: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT:\*  
10: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT:\*  
11: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT:\*  
12: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT:\*  
13: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT:\*  
14: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT:\*  
15: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT:\*  
16: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT:\*  
17: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT:\*  
18: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT:\*  
19: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT:\*  
20: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT:\*  
21: /SID2/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT:\*  
22: /SID2/gcgdata/geneeq/geneeqn-emb1/NA2001.DAT:\*  
23: /SID2/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT:\*  
24: /SID2/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1          | 2277.6 | 97.1        | 2384   | 20    | AA36895     |
| 2          | 2248   | 95.9        | 2407   | 20    | AAV82520    |
| 3          | 2248   | 95.9        | 2407   | 21    | AA62411     |
| 4          | 2150.4 | 91.7        | 2193   | 22    | AAH14279    |
| 5          | 564.8  | 24.1        | 2182   | 23    | ABL01991    |
| 6          | 564.8  | 24.1        | 2539   | 23    | ABL01989    |
| 7          | 551.6  | 23.5        | 620    | 22    | AAH05966    |
| 8          | 517    | 22.0        | 807    | 22    | AAH34971    |
| 9          | 480.6  | 20.5        | 6534   | 23    | ABL01988    |

|    |       |      |         |    |          |                      |
|----|-------|------|---------|----|----------|----------------------|
| 10 | 480.6 | 20.5 | 6541    | 23 | ABL01990 | Drosophila melanog   |
| 11 | 408.6 | 17.4 | 437     | 21 | AAA45541 | Human secreted exp   |
| 12 | 358.6 | 15.3 | 5369    | 20 | AA78080  | A. gossypii ADE4 D   |
| 13 | 316.6 | 13.5 | 1185    | 21 | AA37701  | Arabidopsis thaliana |
| 14 | 299.6 | 12.8 | 2770    | 23 | ABL28787 | Drosophila melanog   |
| 15 | 284.2 | 12.1 | 2984    | 24 | AA27254  | Human transporter    |
| 16 | 282.6 | 12.1 | 1725    | 21 | AA294743 | Human ATP binding    |
| 17 | 282.6 | 12.1 | 1933    | 21 | AA248823 | CDNA encoding a hu   |
| 18 | 282.6 | 12.1 | 1958    | 22 | AA158469 | Human ABC transpor   |
| 19 | 282.6 | 12.1 | 2529    | 22 | AA15604  | Human secreted pro   |
| 20 | 277.6 | 11.8 | 296     | 20 | AA41426  | Human ABCB12 trans   |
| 21 | 274.8 | 11.5 | 2893    | 22 | AA500010 | Drosophila melanog   |
| 22 | 269.2 | 11.5 | 5383    | 23 | ABL28786 | Human polynucleoti   |
| 23 | 255.6 | 10.9 | 1960    | 22 | AA160255 | S. epidermidis ope   |
| 24 | 189.2 | 8.1  | 1647    | 22 | AAH52466 | S. epidermidis gen   |
| 25 | 189.2 | 8.1  | 3039    | 22 | AAH34851 | S. epidermidis gen   |
| 26 | 189.2 | 8.1  | 3585    | 22 | AAH54600 | S. epidermidis gen   |
| 27 | 189.2 | 8.1  | 3736    | 22 | AAH54207 | Staphylococcus epi   |
| 28 | 188.8 | 8.1  | 1749    | 24 | ABN92430 | Staphylococcus aur   |
| 29 | 187.4 | 8.0  | 1053    | 18 | AAV75107 | Bacillus clausii g   |
| 30 | 185.8 | 7.9  | 939     | 24 | ABK79220 | Genomic sequence o   |
| 31 | 185.4 | 7.9  | 2365589 | 24 | ABA90521 | S. aureus polypept   |
| 32 | 184.6 | 7.9  | 1737    | 20 | AA31863  | Staphylococcus aur   |
| 33 | 184.6 | 7.9  | 1737    | 23 | AA554766 | Enterococcus faeca   |
| 34 | 184.4 | 7.8  | 1716    | 23 | AA52885  | Enterococcus faeca   |
| 35 | 180.8 | 7.7  | 1716    | 22 | ABA82943 | Streptococcus pneu   |
| 36 | 180.2 | 7.7  | 7760    | 19 | AAV52196 | Listeria monocytog   |
| 37 | 178.6 | 7.6  | 1424    | 24 | ABQ70460 | Apixid gene. Acti    |
| 38 | 177.8 | 7.6  | 2586    | 24 | ABQ70666 | A. pleuropneumonia   |
| 39 | 170   | 7.2  | 5120    | 18 | AA73218  | Streptococcus pneu   |
| 40 | 170   | 7.2  | 5120    | 21 | AA288585 | E. coli J96 pathog   |
| 41 | 165.8 | 7.1  | 13188   | 19 | AAV52203 | Listeria monocytog   |
| 42 | 165.4 | 7.1  | 3576    | 19 | AAV1285  | Bacillus lichenifo   |
| 43 | 158.4 | 6.8  | 2832    | 24 | ABQ70839 | Enterococcus faeca   |
| 44 | 158.2 | 6.7  | 1233    | 24 | ABK74885 |                      |
| 45 | 158   | 6.7  | 1719    | 22 | ABA82948 |                      |

ALIGNMENTS

RESULT 1  
AA36895 standard; cDNA; 2384 BP.  
AA36895:  
16-JUL-1999 (first entry)  
Human ABC-Transporter-7 (HABCT) coding sequence.  
HABCT; human; ABC-transporter-7; diagnosis; cancer; autoimmune disease;  
Addison's disease; insulin-dependent diabetes mellitus; therapy;  
microsomal disorder; ss.  
Homo sapiens.  
WO9921885-A1.  
06-MAY-1999.  
29-OCT-1997; 97WO-CN00120.  
29-OCT-1997; 97WO-CN00120.  
(UYSH-) UNIV SHANGHAI SECOND MEDICAL.  
Zhang Q;  
WPI; 1999-312945/26.  
P-PSDB; AA14068.  
HABCT a protein useful in the treatment of cancer, diabetes,

antibody bodies

1022

microsomal disorders and Addison's disease  
Claim 2; Page 26-27; 31pp; English.

This sequence encodes the human ABC-transporter-7 (HABC7) protein of the invention. HABC7 is an ABC transporter protein. The DNA, vectors containing it and recombinant host cells are useful for recombinant production of HABC7. The DNA, HABC7 and antibodies against HABC7 are useful as research reagents, for screening assays and in diagnostic assays. Antagonists and agonists of HABC7 can be used to inhibit or enhance, respectively, the activity of HABC7 or its fragments can be coding sequence. Anti-HABC7 antibodies and HABC7 or its fragments are used in vaccines. In particular, the proteins, antibodies, agonists and antagonists can be used for treating, e.g. cancer, autoimmune disease, Addison's disease, microsomal disorders and insulin-dependent diabetes mellitus, related to both an excess and insufficient amounts of HABC7.

Sequence 2384 BP; 719 A; 467 C; 567 G; 631 T; 0 other;

Query Match 97.1%; Score 2277.6; DB 20; Length 2384;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 2318; Conservative 0; Mismatches 24; Indels 6; Gaps 2;

```
QY 1 ATGGGCTGCTCGCGATCATCTTGGCGCGCGCGCGCGCGCTCTTGGAAAAG 60
DB 9 ATGGGCTGCTCGCGATCATCTTGGCGCGCGCGCGCGCGCTCTTGGAAAAG 68
QY 61 CGCGGCACTCGCGATCTGATCGCGCTTTAGTCTCTTTAGCGGCTCAGGTCCGAG 120
DB 69 CGCGGCACTCGCGATCTGATCGCGCTTTAGTCTCTTTAGCGGCTCAGGTCCGAG 128
QY 121 TGGAGGCCATCACTGGCGGCTTTGGGAACGCTCGAGCTACAGATTCAGAGTCA 180
DB 129 TGGAGGCCATCACTGGCGGCTTTGGGAACGCTCGAGCTACAGATTCAGAGTCA 188
QY 181 TTAAGAGATCACTGAGAGATTTGGGAAAGCAATTCAGACAAGTTCTTAGATGCT 240
DB 189 TTTAAAGATCACTGAGAGATTTGGGAAAGCAATTCAGACAAGTTCTTAGATGCT 248
QY 241 GCAAGGCTCTCGAGTATGGCCACTGATGAAAAAGAGACATGTTGGATGTCATGCA 300
DB 249 GCAAGGCTCTCGAGTATGGCCACTGATGAAAAAGAGACATGTTGGATGTCATGCA 308
QY 301 GGAGAGGCTCTCAACACACACCCAAAGAGGTTAAAGATGTTGATCTCGAAAAATC 360
DB 309 GGAGAGGCTCTCAACACACACCCAAAGAGGTTAAAGATGTTGATCTCGAAAAATC 368
QY 361 ATAAAAGCAATGCTTTCTATGTTGGCCCAAGACAGGCCAGATCTACAGACTAGATT 420
DB 369 ATAAAAGCAATGCTTTCTATGTTGGCCCAAGACAGGCCAGATCTACAGACTAGATT 428
QY 421 GCGATTTGCTGGGATTTTGGTGGTGCMAAGGCCATTAATTTGGTTCCTTCATG 480
DB 429 CCCATTTGCTGGGATTTTGGTGGTGCMAAGGCCATTAATTTGGTTCCTTCATG 488
QY 481 TTTAAATATGCTGTAGACAGCCTCAACAGATGTCGGAAGAAACATGCTAGATGAT 540
DB 489 TTTAAATATGCTGTAGACAGCCTCAACAGATGTCGGAAGAAACATGCTAGATGAT 548
QY 541 GCACCAATATACAGTTGCAACCATGCAACAGCAGTTCTGATTTGGCTATGCTATCAAG 600
DB 549 GCACCAATATACAGTTGCAACCATGCAACAGCAGTTCTGATTTGGCTATGCTATCAAG 608
QY 601 GCTGAGCTGCTTTTAAACGAAGTTGCAATGCTATTTGGCAAGTATGCCGAGAT 660
DB 609 GCTGAGCTGCTTTTAAACGAAGTTGCAATGCTATTTGGCAAGTATGCCGAGAT 668
QY 661 TCAATCCGAAGATAGCAAAATGCTTTCTCATCTTCAACAACCTGATCTGGGTTTT 720
DB 669 TCAATCCGAAGATAGCAAAATGCTTTCTCATCTTCAACAACCTGATCTGGGTTTT 728
QY 721 CACCTGAGACAGACAGGAGCTTTATCTAAGCTATTGACAGAGAACAGGGGTATCT 780
```

```
DB 729 CACCTGAGACAGACAGGAGCTTTATCTAAGCTATTGACAGAGAACAGGGGTATCT 788
QY 781 AGTTTGCTGCTGAGTCTTTGGATTTAATCTTCTCCCATCATGTTTGAAGTATG--- 837
DB 789 AGTTTGCTGCTGAGTCTTTGGATTTAATCTTCTCCCATCATGTTTGAAGTATGCTT 848
QY 838 CTGTGAGTGTGTTTGTATTAACAATGCGGTGCCAGTTGCTTTGGTATCCCTTGA 897
DB 849 CTGTGAGTGTGTTTGTATTAACAATGCTGTGCCAGTT---GCTGTGAACCTTGA 905
QY 898 ACACTTGTATCAACACAGATTCACAGTTGCAACAGGAGGAGAACTAGATTAGA 957
DB 906 ACACTTGTATCAACACAGATTCACAGTTGCAACAGGAGGAGAACTAGATTAGA 965
QY 958 ATGAATGAACAAAGAGATATGATGCAAGTATGCTGTATGACTCTGCTGAAT 1017
DB 966 TTGGAATTTAACAAGAGATTAATGATGCAAGTATGCTGTATGACTCTGCTGAAT 1025
QY 1018 TATGAACCTGTGAAGTATTTTAAATGAAGATATGAACACAGATTTGATGATTT 1077
DB 1026 TATGAACCTGTGAAGTATTTTAAATGAAGATATGAACACAGATTTGATGATTT 1085
QY 1078 TTGAAGACATATGAGACTGCTTCAATGAAGAAAGTACCTTACCTGGCTATGCTGA 1137
DB 1086 TTGAAGACATATGAGACTGCTTCAATGAAGAAAGTACCTTACCTGGCTATGCTGA 1145
QY 1138 GGTCAAGTCTATTTTCACTGTCGTTTAAACAGCTATTAATGCTGCTCGCACTCAGGA 1197
DB 1146 GGTCAAGTCTATTTTCACTGTCGTTTAAACAGCTATTAATGCTGCTCGCACTCAGGA 1205
QY 1198 ATTGTGCAAGTACCTTACTGTTGAGATCTAGTAATGCTGAATGGAACGCTTTTCA 1257
DB 1206 ATTGTGCAAGTACCTTACTGTTGAGATCTAGTAATGCTGAATGGAACGCTTTTCA 1265
QY 1258 CTTTCAATACCCCGAACTTTTGGGAACTGTATATGAGAGACTAGACAAGCACTATA 1317
DB 1266 CTTTCAATACCCCGAACTTTTGGGAACTGTATATGAGAGACTAGACAAGCACTATA 1325
QY 1318 GATATGAACACCTTGTACTCTACTCAAGGTAGACACCCAAATTAAGACAAAGTATG 1377
DB 1326 GATATGAACACCTTGTACTCTACTCAAGGTAGACACCCAAATTAAGACAAAGTATG 1385
QY 1378 GCATCTCCCTCTTCAAGATCACACACAGACAGTACCGTGGCTTTGATATGTCATTT 1437
DB 1386 GCATCTCCCTCTTCAAGATCACACACAGACAGTACCGTGGCTTTGATATGTCATTT 1445
QY 1438 GAATPACATTTGAGGCGCAAGAAAGTCTTATGAGAAATCTTTGAAGTCCCTGACAGAA 1497
DB 1446 GAATPACATTTGAGGCGCAAGAAAGTCTTATGAGAAATCTTTGAAGTCCCTGACAGAA 1505
QY 1498 AAAGTGGCATTTGTAGAGATGAGGTCAAGGAAAGCACAATAGTGAAGCTATTAAT 1557
DB 1506 AAAGTGGCATTTGTAGAGATGAGGTCAAGGAAAGCACAATAGTGAAGCTATTAAT 1565
QY 1558 CGCTTTATAGGCTCAAAAAGGTAGCATTTATCTGTGTGCAAAATATATCAAGAGTG 1617
DB 1566 CGCTTTATAGGCTCAAAAAGGTAGCATTTATCTGTGTGCAAAATATATCAAGAGTG 1625
QY 1618 AGCCTGAAGGCTTCGAGGAGGAGTGGTATGCTCAGATGCTGTCTCTTCAT 1677
DB 1626 AGCCTGAAGGCTTCGAGGAGGAGTGGTATGCTCAGATGCTGTCTCTTCAT 1685
QY 1678 AATACATTTATTAACAACCTCTTATATGAAACATCAGTGTCTTCACTGAGAAAGTAT 1737
DB 1686 AATACATTTATTAACAACCTCTTATATGAAACATCAGTGTCTTCACTGAGAAAGTAT 1745
QY 1738 GCAAGTGGCAAAATAGCTGCACTTATATGCAATTTCTTGAATGCCACATGAGATATAC 1797
DB 1746 GCAAGTGGCAAAATAGCTGCACTTATATGCAATTTCTTGAATGCCACATGAGATATAC 1805
QY 1799 ACCCAAGTAGGGAACGAGGACTCAAGCTTTCAGAGAGAGAAAGCAAGATGACAAAT 1857
DB 1806 ACCCAAGTAGGGAACGAGGACTCAAGCTTTCAGAGAGAGAAAGCAAGATGACAAAT 1865
```





```

Db 721 GCAGACACGCGGAGCTTATCTAAGGCTATGACAGAGAAACAGGGTATCACTTTTG 780
Qy 788 TCCCGAGGCTTTGGATTTTATCTTCCATCATGTTGAAGTGAAGCTGGCAGTG 847
Db 781 TCCGAGGCTTTGGATTTTATCTTCCATCATGTTGAAGTGAAGCTGGCAGTG 840
Qy 848 GTGTTTGTATTAACAATGCGGTGCCAGTTTGTGGTAAACCTTGGAACTTTGTA 907
Db 841 GTGTTTGTATTAACAATGCGGTGCCAGTTTGTGGTAAACCTTGGAACTTTGTA 900
Qy 908 CATACACAGCATTCACAGTTGCACTACACGCTGAGAACTTGAATTAAGAAATGA 967
Db 901 CATACACAGCATTCACAGTTGCACTACACGCTGAGAACTTGAATTAAGAAATGA 960
Qy 968 ACAAGGAGATATGATGAGGTAATGCTGCTATAGACTCATGCTGAATTAAGAACTG 1027
Db 961 ACAAGGAGATATGATGAGGTAATGCTGCTATAGACTCATGCTGAATTAAGAACTG 1020
Qy 1028 TGAAGTATTTTATATGAAAGATATGAAGACAGAGATATGAGATTTTGAAGAGT 1087
Db 1021 TGAAGTATTTTATATGAAAGATATGAAGACAGAGATATGAGATTTTGAAGAGT 1080
Qy 1088 ATGAGACTGCTTCATTGAAAAAGTACCTCTACTCTGGCTATGCTGAACCTTGGTCAAAGTG 1147
Db 1081 ATGAGACTGCTTCATTGAAAAAGTACCTCTACTCTGGCTATGCTGAACCTTGGTCAAAGTG 1140
Qy 1148 CTATTTTCAAGTGTGGTTTAAACAGCTATTAATGCTGCTGCCAGTCAAGGAAATTTGGCAG 1207
Db 1141 CTATTTTCAAGTGTGGTTTAAACAGCTATTAATGCTGCTGCCAGTCAAGGAAATTTGGCAG 1200
Qy 1208 GTACCTTACTGTTGGAAGATCTAGTAAATGTAATGTAAGTGCCTTTTCAAGCTTTTCATTAC 1267
Db 1201 GTACCTTACTGTTGGAAGATCTAGTAAATGTAATGTAAGTGCCTTTTCAAGCTTTTCATTAC 1260
Qy 1268 CCTGAACTTTCTGAGAACTGTATATAGAGAGACTGAAACAGCACTCATGATATGAACA 1327
Db 1261 CCTGAACTTTCTGAGAACTGTATATAGAGAGACTGAAACAGCACTCATGATATGAACA 1320
Qy 1328 CCTGTTTACTGTTCAAGGTAAGACACCCAAATTAAGAAAGTAAGTGCATCTCCC 1387
Db 1321 CCTGTTTACTGTTCAAGGTAAGACACCCAAATTAAGAAAGTAAGTGCATCTCCC 1380
Qy 1388 TTCAGATCAACACAGACAGCTACCGTGGCCTTGTATATGATTTTGAATCATTTG 1447
Db 1381 TTCAGATCAACACAGACAGCTACCGTGGCCTTGTATATGATTTTGAATCATTTG 1440
Qy 1448 AGGGCCAGAAAGTCTTAAAGTGAATATCTTTGAAGTCCCTGACAGAAAGAAAGTGCCCA 1507
Db 1441 AGGGCCAGAAAGTCTTAAAGTGAATATCTTTGAAGTCCCTGACAGAAAGAAAGTGCCCA 1500
Qy 1508 TTTGTAAGAGTATGTTGGTCAAGGAAAGCAAAATAGTGAAGCTATTTTGGCTTCTATG 1567
Db 1501 TTTGTAAGAGTATGTTGGTCAAGGAAAGCAAAATAGTGAAGCTATTTTGGCTTCTATG 1560
Qy 1568 AGCCTCAAAAGGGTGAAGCTTATCTTGTGCTGCTCAAAATTAACAAGATGAGACCTGAAA 1627
Db 1561 AGCCTCAAAAGGGTGAAGCTTATCTTGTGCTGCTCAAAATTAACAAGATGAGACCTGAAA 1620
Qy 1628 GCTTTGGAAGGAGTGGAGTGTACCTCAGAGATGCTGCTCTTCCATATATCTATTT 1687
Db 1621 GCTTTGGAAGGAGTGGAGTGTACCTCAGAGATGCTGCTCTTCCATATATCTATTT 1680
Qy 1688 ATTACAACTCTTATATGAAAACATCAGTCTTCACTGAGAGAGTATGAGAGTGCGCA 1747
Db 1681 ATTACAACTCTTATATGAAAACATCAGTCTTCACTGAGAGAGTATGAGAGTGCGCA 1740
Qy 1748 AATTAGCTGACTTATGATGCAATTTCTGCAATGCAACAGATATGCAACCAAGTAG 1807
Db 1741 AATTAGCTGACTTATGATGCAATTTCTGCAATGCAACAGATATGCAACCAAGTAG 1800
Qy 1808 GGGAAAGAGACTCAAGCTTTGAGAGAGAAAGCAAAAGTATGCAATTTGCAAGAGCA 1867

```

```

Db 1801 GGGAAAGAGACTCAAGCTTTGAGAGAGAAAGCAAAAGTATGCAATTTGCAAGAGCA 1860
Qy 1868 TTTTGAAGAGACCCCGCATCTATCTATGATGAAGCTACTTCATCGTTAGATTCGATTA 1927
Db 1861 TTTTGAAGAGACCCCGCATCTATCTATGATGAAGCTACTTCATCGTTAGATTCGATTA 1920
Qy 1928 CTGAAGAGACTATTTCTTGTGTCATGAAGATGTGTCAAAACAGAACTTTTATTTTCA 1987
Db 1921 CTGAAGAGACTATTTCTTGTGTCATGAAGATGTGTCAAAACAGAACTTTCTATTTTCA 1980
Qy 1988 TTGCAACACAGTATGTCACACAGTGTGTATGACAGATGAATTAATTTGTTGATTCAGGTA 2047
Db 1981 TTGCAACACAGTATGTCACACAGTGTGTATGACAGATGAATTAATTTGTTGATTCAGGTA 2040
Qy 2048 AGTAGCCGAACGTGTATACCAACCATGATGCTTGTGCTTAAACCTCATATATCTATTCA 2107
Db 2041 AGTAGCCGAACGTGTATACCAACCATGATGCTTGTGCTTAAACCTCATATATCTATTCA 2100
Qy 2108 AATGTGGCATACACAGAGCAGCGGTGCGAACCATGATMACCCCAATGGAGCA 2167
Db 2101 AATGTGGCATACACAGAGCAGCGGTGCGAACCATGATMACCCCAATGGAGCA 2160
Qy 2168 AGAAAGAAATATATCCAAAGAGAGAAAGAAAGAACTTCAAGAAAGAAATTTGCAATA 2227
Db 2161 AGAAAGAAATATATCCAAAGAGAGAAAGAAAGAAAGAACTTCAAGAAAGAAATTTGCAATA 2220
Qy 2228 GTGTGAAGAGGCTGTGGAAGCTGTGCTGCTAGTACATATGAGATTTTCTTTTGT 2286
Db 2221 GTGTGAAGAGGCTGTGGAAGCTGTGCTGCTAGTACATATGAGATTTCTTTTGT 2278

```

```

RESULT 3
AAA62411
ID AAA62411 standard; cDNA; 2407 BP.
XX
AC AAA62411;
XX
DT 13-NOV-2000 (first entry)
XX
DE Human ATP-binding cassette transport protein nucleotide sequence.
KW Human; ATP-binding cassette transport protein; ABCxH; cytosolic;
KW nootropic; neuroprotective; cerebroprotective; antidepressant;
KW anticonvulsant; antiparkinsonian; neuroleptic; cancer; epilepsy;
KW neurodegenerative disorder; akathisia; amorphous lateral sclerosis;
KW bipolar disorder; cataplexy; dementia; depression; Down's syndrome;
KW tardive dyskinesia; dystonia; multiple sclerosis; neurofibromatosis;
KW schizophrenia; Tourette's disease; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 9..2252
FT /lag= a
FT /product= "ABCxH"
XX
PN US6080842-A.
XX
PD 27-JUN-2000.
XX
PD 18-NOV-1998; 98US-0195391.
XX
PR 17-JUL-1997; 97US-0895522.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Shah P, Corley NC, Hillman JL;
XX
DR WPI: 2000-451228/39.
DR P-PSDB; AAB13355.
XX
PT New human ATP-binding cassette transport protein useful for diagnosing,
PT preventing and treating cancers, e.g. leukemia, sarcoma, and neuronal

```

PT disorders, e.g. schizophrenia, dementia -  
XX  
PS Example 5; Fig 1; 31pp; English.

Example 5; Fig 1; 31pp; English.

CC The present sequence encodes a human ATP-binding cassette transport  
CC protein (ABCTXH). The nucleotide sequence was isolated from a human  
CC ovarian tissue cDNA library. Clones from the library were sequenced and  
CC used as query sequences against homology databases. ABCTXH encoding  
CC polynucleotides were extended using PCR. ABCTXH polynucleotides and  
CC polypeptides may be used for the diagnosis, prevention and treatment of  
CC conditions or disorders associated with the expression of ABCTXH, e.g.  
CC adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, akathisia, Alzheimer's disease, amnesia, amyotrophic  
CC lateral sclerosis, bipolar disorder, cataplexy, cerebral neoplasms,  
CC dementia, depression, Down's syndrome, tardive dyskinesia, dystonias,  
CC epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis,  
CC Parkinson's disease, paranoid psychoses, schizophrenia, and Tourette's  
CC disease.

**SQ** Sequence 2407 BP; 726 A; 477 C; 570 G; 629 T; 5 other;

Query Match 95.9%; Score 2248; DB 21; Length 2407;

Best Local Similarity 99.4%; Pred. No. 0;

Matches 2265; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

|    |     |   |     |
|----|-----|---|-----|
| Qy | 8   | TGCGGGATGCAATCTTTGGCGGTGGGGCGCGCCGGGGCGCTGTTTGGAAAAGCCCGCGC   | 67  |
| Db | 1   | TGCTGGCATGCAATCTTTGGCGGTGGGGCTGCGGGGGCGTGTCTTTGGAAAAGCCCGCGC  | 60  |
| Qy | 68  | ACTCCGGCATCTGATCCGAGCTTTAGTCTGTGTACGGGTGAGGTCCGACATGGAGGC     | 127 |
| Db | 61  | ACTCCGGATTCATATCCGAGCTTTAGTCTGTGTACGGGTGAGGTCCGACATGGAGGC     | 120 |
| Qy | 128 | CACATCAACTCGGGCGCTTGGGAAACGCTCGAGCTACAGATTTCCAGATCTTAAAA      | 187 |
| Db | 121 | CACATCAACTCGGGCGCTTGGGAAACCGGTGAGCTACAGATTTCCAGATCTTAAAA      | 180 |
| Qy | 188 | GTATATCACTGGCAGAGATTTGGGAAAAGGCAATTGAGACAGTTCTTAGATGTCGAAAG   | 247 |
| Db | 181 | GTATATCACTGGCAGAGATTTGGGAAAAGGCAATTGAGACAGTTCTTAGATGTCGAAAG   | 240 |
| Qy | 248 | CTCTCCAGGTATGGCCACTGATATAAAAAGGACATGTGGCANTGGCATGACGAGGAG     | 307 |
| Db | 241 | CTCTCCAGGTATGGCCACTGATATAAAAAGGACATGTGGCANTGGCATGACGAGGAG     | 300 |
| Qy | 308 | GACTCCACACAGACCCCAAAAAGAGGTTTAAAAGATGTTGATCTCGGAAAATCATAAAG   | 367 |
| Db | 301 | GACTCCACACAGACCCCAAAAAGAGGTTTAAAAGATGTTGATCTCGGAAAATCATAAAG   | 360 |
| Qy | 368 | CAATGCTTTCTTATGTGTGGCCCAAGACAGGCCAGATCTACGAGCTAGAGTTGCCATTT   | 427 |
| Db | 361 | CAATGCTTTCTTATGTGTGGCCCAAGACAGGCCAGATCTACGAGCTAGAGTTGCCATTT   | 420 |
| Qy | 428 | CGCTGGGATTTTGGGTGGTGGCATTGATATTGGTGTCCCTTATGTTTAAAT           | 487 |
| Db | 421 | CGCTGGGATTTTGGGTGGTGGCATTGATATTGGTGTCCCTTATGTTTAAAT           | 480 |
| Qy | 488 | ATGCTGTAGACAGCTCAACACAGATGTGGGAAAACATGCTGAACCTGAGTAGACCA      | 547 |
| Db | 481 | ATGCTGTAGACAGCTCAACACAGATGTGGGAAAACATGCTGAACCTGAGTAGACCA      | 540 |
| Qy | 548 | ATACAGTTGCACATCCATGGCAACAGCATTTGATTTGGCTATGGTGTCAAGAGCTGGAG   | 607 |
| Db | 541 | ATACAGTTGCACATCCATGGCAACAGCATTTGATTTGGCTATGGTGTATCAAGAGCTGGAG | 600 |
| Qy | 608 | CTGCTTTTTTTAACGAGATTGAAATGCAATTTGGCAAGGTAGCCCAATTCATTC        | 667 |
| Db | 601 | CTGCTTTTTTTAACGAGATTGAAATGCAATTTGGCAAGGTAGCCCAATTCATTC        | 660 |
| Qy | 668 | GAAACAATAGCCAAAATGTCTTTCTCATCTTACACAACCTGGAATCTGGTTCACCTGA    | 727 |
| Db | 661 | GAAACAATAGCCAAAATGTCTTTCTCATCTTACACAACCTGGAATCTGGTTCACCTGA    | 720 |

|    |      |  |     |
|----|------|--|-----|
| QY | 728  | GCAGCAGACGGGAGCTTTATCTTAAGGCTATTGACAGAGAAACAAAGGGATTCAGTTTTG | 787 |
| Db | 721  | GCAGCAGACGGGAGCTTTATCTTAAGGCTATTGACAGAGAAACAAAGGGATTCAGTTTTG | 780 |
| QY | 788  | TCTCGAGGCTTGGTATTTAAATCTTCTCCACCAATGTTGAAGTGAAGCTGTGACGTG    | 847 |
| Db | 781  | TCTCGAGGCTTGGTATTTAAATCTTCTCCACCAATGTTGAAGTGAAGCTGTGACGTG    | 840 |
| QY | 848  | GTGTTTGTATTACAAATGCGGTGCCAGTTTGCTTGTGAACCTTGGAAACCTTGGTA     | 907 |
| Db | 841  | GTGTTTGTATTACAAATGCGGTGCCAGTTTGCTTGTGAACCTTGGAAACCTTGGTA     | 900 |
| QY | 908  | CATACACAGCACTTACACGTTGCGACTCACACGGTGGAGAACTAGATTAGAAATGA     | 967 |
| Db | 901  | CATACACAGCACTTACACGTTGCGACTCACACGGTGGAGAACTAGATTAGAAATGA     | 960 |
| QY | 968  | ACAAAGCAGATPAATGATGACAGGTAAATGCTGTCTATGACTACGCTGAATTTGAAC    | 102 |
| Db | 961  | ACAAAGCAGATPAATGATGACAGGTAAATGCTGTCTATGACTACGCTGAATTTGAAC    | 102 |
| QY | 1028 | TGAAGTATTTTAAATGAAGATATGAAACAAGAGATATGATGATTTTGAAGACGT       | 108 |
| Db | 1021 | TGAAGTATTTTAAATGAAGATATGAAACAAGAGATATGATGATTTTGAAGACGT       | 108 |
| QY | 1088 | ATGAGACTGCTTCACTTGAAGAAATGCTCTATCTGTGCTATGCTGAACCTTGGTCA     | 114 |
| Db | 1081 | ATGAGACTGCTTCACTTGAAGAAATGCTCTATCTGTGCTATGCTGAACCTTGGTCA     | 114 |
| QY | 1148 | CTATTTTCAGTGTGGTTTAAACGCTATAATGCTGTCCGACGTCAAGGAAATTTGGG     | 120 |
| Db | 1141 | CTATTTTCAGTGTGGTTTAAACGCTATAATGCTGTCCGACGTCAAGGAAATTTGGG     | 120 |
| QY | 1208 | GTACCTTTACTGTGAGATCTAGTAAATGCTGAATGGAAGCTCTTTTCACTTCACTAC    | 126 |
| Db | 1201 | GTACCTTTACTGTGAGATCTAGTAAATGCTGAATGGAAGCTCTTTTCACTTCACTAC    | 126 |
| QY | 1268 | CCCTGAACCTTTCTGGGAACGTATATATAGAGAGCTAGACAAGCACTCATATGATGAA   | 132 |
| Db | 1261 | CCCTGAACCTTTCTGGGAACGTATATATAGAGAGCTAGACAAGCACTCATATGATGAA   | 132 |
| QY | 1328 | CCTTGTTTACTCTACTCAAGGTAGACACCCAAATTAAACAAAGTATGGCATCTCCCC    | 138 |
| Db | 1321 | CCTTGTTTACTCTACTCAAGGTAGACACCCAAATTAAACAAAGTATGGCATCTCCCC    | 138 |
| QY | 1388 | TTCAATATCACACACAGACAGCTACCGTGGCCTTTGATATGTCATTTTGAATACAT     | 144 |
| Db | 1381 | TTCAATATCACACACAGACAGCTACCGTGGCCTTTGATATGTCATTTTGAATACAT     | 144 |
| QY | 1448 | AGGGCCAGAAAGTCTTATGTAATATCCTTTGAAGTCCCTGACAGGAAAGAAATGG      | 150 |
| Db | 1441 | AGGGCCAGAAAGTCTTATGTAATATCCTTTGAAGTCCCTGACAGGAAAGAAATGG      | 150 |
| QY | 1508 | TTTATGAGAGGTATGAGGTCAAGGAAAGACAAATAGTAGGCTATTAATTTGCTT       | 156 |
| Db | 1501 | TTTATGAGAGGTATGAGGTCAAGGAAAGACAAATAGTAGGCTATTAATTTGCTT       | 156 |
| QY | 1568 | AGGCTCAAAAGGGTATGATTTATCTTGCTGTGCTCAAAATATACAAAGTATGAGCT     | 162 |
| Db | 1561 | AGGCTCAAAAGGGTATGATTTATCTTGCTGTGCTCAAAATATACAAAGTATGAGCT     | 162 |
| QY | 1628 | GCTTTGGAAGGAGTGGGAGTGTACTTCAGAGATGCTGTCTTCCATTAATACTAT       | 168 |
| Db | 1621 | GCTTTGGAAGGAGTGGGAGTGTACTTCAGAGATGCTGTCTTCCATTAATACTAT       | 168 |
| QY | 1688 | ATTACAACTCTTATATGAAACATCAATGCTTCACTTGAAGATGTATGCAATGG        | 174 |
| Db | 1681 | ATTACAACTCTTATATGAAACATCAATGCTTCACTTGAAGATGTATGCAATGG        | 174 |
| QY | 1748 | AATTAAGCTGACCTTCAATGATGCAATCTTGGAATGCAATGGATATGCAACCAAG      | 180 |
| Db | 1741 | AATTAAGCTGACCTTCAATGATGCAATCTTGGAATGCAATGGATATGCAACCAAG      | 180 |
| QY | 1808 | GGGAGACAGACTCAAGCTTTCAGAGAGAGAAAGCAAGATGACCAATTTGCACAG       | 186 |

```
|||||
Db 1801 GGAACGAGACTCAAGCTTTGAGAGAGGAGAAAGCAAGAGTACCAATTGCAAGAGCCA 1860
Qy 1868 TTTTGAAGAGACCCCGACGATCACTTATGATGAAGACTCTTCATCGTTAGATTGATTA 1927
Db 1861 TTTTGAAGAGACCCCGACGATCACTTATGATGAAGACTCTTCATCGTTAGATTGATTA 1920
Qy 1928 CTGAAGAGACTATCTTGGTGCATGAAGAGATGGTGGTCAAGACAGAACTTCTATTTC 1987
Db 1921 CTGAAGAGACTATCTTGGTGCATGAAGAGATGGTGGTCAAGACAGAACTTCTATTTC 1980
Qy 1988 TTGCACACAGATTTGTCAACAGTGGTTGATGACAGATGAATCATTTGTTGATCAGGGTA 2047
Db 1981 TTGCACACAGATTTGTCAACAGTGGTTGATGACAGATGAATCATTTGTTGATCAGGGTA 2040
Qy 2048 AGGTGCGGCAAGTGGTACCCGACGATGTTGCTTGTGCTTACCCCTCATGATCTATTTCAG 2107
Db 2041 AGGTGCGGCAAGTGGTACCCGACGATGTTGCTTGTGCTTACCCCTCATGATCTATTTCAG 2100
Qy 2108 AATGTGGCATACACAGAGCAGCCGCTGTGTCAGAACCATGATATACCCCAATGGAGACA 2167
Db 2101 AATGTGGCATACACAGAGCAGCCGCTGTGTCAGAACCATGATATACCCCAATGGAGACA 2160
Qy 2168 AGAAGAAATATATATCCAAAGAGAGAAAGAAAGAACTACAGAGAAATTTGCAATA 2227
Db 2161 AGAAGAAATATATATCCAAAGAGAGAAAGAAAGAACTACAGAGAAATTTGCAATA 2220
Qy 2228 GTGTGAAAAGCTGTGGAAAATCTGTGCTAGTCAATTAACATTTCTTTTTCG 2286
Db 2221 GTGTGAAAAGCTGTGGAAAATCTGTGCTAGTCAATTAACATTTCTTTTTCG 2278

RESULT 4
AAH14279
ID AAH14279 standard; cDNA; 2193 BP.
AC AAH14279;
XX
XX
DT 26-JUN-2001 (first entry)
DE Human cDNA sequence SEQ ID NO:11609.
XX
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX
OS Homo sapiens.
XX
XX
EN EP1074617-A2.
XX
XX
PD 07-FEB-2001.
XX
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX
PA (HELI-) HELIX RES INST.
XX
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX
WP1; 2001-318749/34.
XX
XX
PR Primer sets for synthesizing polynucleotides, particularly the 5602
PR full-length cDNAs defined in the specification, and for the detection
PR and/or diagnosis of the abnormality of the proteins encoded by the
PR full-length cDNAs -
XX
XX
PS Claim 8; SEQ ID 11609; 2537bp + CD ROM; English.
XX
XX
CC The present invention describes primer sets for synthesizing 5602
```

```
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX
SQ Sequence 2193 BP; 621 A; 451 C; 536 G; 585 T; 0 other;
Query Match 91.7%; Score 2150.4; DB 22; Length 2193;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2164; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
Qy 1 ATGGCGTCTGCTCGGATGCAATTTCTGGCGGCGCGCGCGGCGCTTGGAAAAG 60
Db 26 ATGGCGTCTGCTCGGATGCAATTTCTGGCGGCGCGCGCGGCGCTTGGAAAAG 85
Qy 61 GCGCGGCACTCCGCGATCTGATCGGCGCTTATGCTCTGTTAGCGGCTCCGAG 120
Db 86 GCGCGGCACTCCGCGATCTGATCGGCGCTTATGCTCTGTTAGCGGCTCCGAG 145
Qy 121 TGAAGGCCATCACTCGGCGCCTTGGGAAACGCTCGAGCCTAC--CAGATTCAGAG 177
Db 146 TGAAGGCCATCACTCGGCGCCTTGGGAAACGCTCGAGCCTACAGACAGATTCAGAG 205
Qy 178 TCATTTAAAAGTATCACTGCGAGAGATTGGGAAAAGGCAATTCAGACAGTCTTTAGAT 237
Db 206 TCATTTAAAAGTATCACTGCGAGAGATTGGGAAAAGGCAATTCAGACAGTCTTTAGAT 265
Qy 238 GCTGCAAGGCTCTCCAGGTATGGCACTGATAGAAAAGAGACATGTTGCGATGTCAT 297
Db 266 GCTGCAAGGCTCTCCAGGTATGGCACTGATAGAAAAGAGACATGTTGCGATGTCAT 325
Qy 298 GCAGAGAGAGACTCCACACAGACCCAAAGAGGTTAAAAGATGTTGATATCCGAAA 357
Db 326 GCAGAGAGAGACTCCACACAGACCCAAAGAGGTTAAAAGATGTTGATATCCGAAA 385
Qy 358 ATCATTAAGCAATGCTTTCTTATGTGTGGCCCAAAGACAGCCGATCTACAGCTTGA 417
Db 386 ATCATTAAGCAATGCTTTCTTATGTGTGGCCCAAAGACAGCCGATCTACAGCTTGA 445
Qy 418 GTTGCCATTGCGTGGGATTTTGGGTGTCGAAAGGCCATGAATATTTGTTCCCTTC 477
Db 446 GTTGCCATTGCGTGGGATTTTGGGTGTCGAAAGGCCATGAATATTTGTTCCCTTC 505
Qy 478 ATGTTTAAATATGCTGTAGACAGGCTCAACAGATGTGCGGAAAATGCTGAACCTGAGT 537
Db 506 ATGTTTAAATATGCTGTAGACAGGCTCAACAGATGTGCGGAAAATGCTGAACCTGAGT 565
Qy 538 GATGCAACCAATATACGTTGCAACATGCAACAGAGTTCTGATTTGCTATGCTATCA 597
Db 566 GATGCAACCAATATACGTTGCAACATGCAACAGAGTTCTGATTTGCTATGCTATCA 625
Qy 598 AGAGCTGAGAGCTGCTTTTAAAGAAAGTGAATGAGATTTGGCAAGTGGCCAG 657
Db 626 AGAGCTGAGAGCTGCTTTTAAAGAAAGTGAATGAGATTTGGCAAGTGGCCAG 685
Qy 658 AATTCATCCGAGAAATAGCCAAAATGTCTTCCATCTTCAACACTGATCTGGGT 717
```

Db 686 AATTCATCCAGAGATAGCCAAAATATGCTTCCATCTTCAACATCTGATCTGGGT 745  
 Qy 718 TTTCACCTGACAGACAGACCGGAGCTTTATCTAAGCTATGACAGAGAAAGGGGT 777  
 Db 746 TTTCACCTGACAGACAGCGGAGCTTTATCTAAGCTATGACAGAGAAAGGGGT 805  
 Qy 778 ATCAGTTTGTCTGAGTGTCTTTGGTATTAATCTTCTCCATCAGTTTGAAGTAG 837  
 Db 806 ATCAGTTTGTCTGAGTGTCTTTGGTATTAATCTTCTCCATCAGTTTGAAGTAG 865  
 Qy 838 CTGTGACGTGGTGTGTTGATTAACAATGCGGTGCCAGTTTGTCTTGGTAAACCTTGA 897  
 Db 866 CTGTGACGTGGTGTGTTGATTAACAATGCGGTGCCAGTTTGTCTTGGTAAACCTTGA 925  
 Qy 898 ACACCTGTACATACACAGCATTCACAGTTGACATCAGACCGGTGAGAGATTAAGA 957  
 Db 926 ACACCTGTATATACACAGCATTCACAGTTGACATCAGACCGGTGAGAGATTAAGA 985  
 Qy 958 ATGAAATGAACAAAGCAGATTAATGATGACAGTAACTGCTATAGACTCATCTGTAAT 1017  
 Db 986 ATGAAATGAACAAAGCAGATTAATGATGACAGTAACTGCTATAGACTCATCTGTAAT 1045  
 Qy 1018 TATGAACTGTGAAGTATTTATATGAAGATATGAAGACAGAGATTAATGATGATTT 1077  
 Db 1046 TATGAACTGTGAAGTATTTATATGAAGATATGAAGACAGAGATTAATGATGATTT 1105  
 Qy 1078 TTGAAGAGTATGAGACGTCTTCACTGTAAGAACTGCTTCACTGCGGTATGCTGAACCTT 1137  
 Db 1106 TTGAAGAGTATGAGACGTCTTCACTGTAAGAACTGCTTCACTGCGGTATGCTGAACCTT 1165  
 Qy 1138 GGTCAAAAGTGTATTTTCACTGTCGATTTAAACAGTAAATGATGCTGCGGACAGTCAAG 1197  
 Db 1166 GGTCAAAAGTGTATTTTCACTGTCGATTTAAACAGTAAATGATGCTGCGGACAGTCAAG 1225  
 Qy 1198 ATTGTGGCAGGTATCCCTTACCTGTTGGAGATTAATGATGATGATGATGATGATGATG 1257  
 Db 1226 ATTGTGGCAGGTATCCCTTACCTGTTGGAGATTAATGATGATGATGATGATGATGATG 1285  
 Qy 1258 CTTCATTAATCCCTGACCTTCTGGAACCTGATATATGAGAGATTAAGACAGACCTGATA 1317  
 Db 1286 CTTCATTAATCCCTGACCTTCTGGAACCTGATATATGAGAGATTAAGACAGACCTGATA 1345  
 Qy 1318 GATATGAACACCTTGTACTCTACTCAAGGTAGACACCAATTAAGAACAAGTAGT 1377  
 Db 1346 GATATGAACACCTTGTACTCTACTCAAGGTAGACACCAATTAAGAACAAGTAGT 1405  
 Qy 1378 GCATCTCCCTTCAAGTACACACAGACAGCTACCGTGGCTTTGATTAATGTCATTTT 1437  
 Db 1406 GCATCTCCCTTCAAGTACACACAGACAGCTACCGTGGCTTTGATTAATGTCATTTT 1465  
 Qy 1438 GAATACATTTGAGGCGCAAGAAAGTCTTAGTGATATCTTTGAAGTCCCTGACGGAAG 1497  
 Db 1466 GAATACATTTGAGGCGCAAGAAAGTCTTAGTGATATCTTTGAAGTCCCTGACGGAAG 1525  
 Qy 1498 AAAGTGGCATTGTAGAGAGTGTGCTCAGGAAAGACAAATAGTGAAGCTTATTTT 1557  
 Db 1526 AAAGTGGCATTGTAGAGAGTGTGCTCAGGAAAGACAAATAGTGAAGCTTATTTT 1585  
 Qy 1558 CGCTTCTATGAGCTCAAAAAGGTAGCATTTATCTTGTGCTCAAAATATATCAAGATGTG 1617  
 Db 1586 CGCTTCTATGAGCTCAAAAAGGTAGCATTTATCTTGTGCTCAAAATATATCAAGATGTG 1645  
 Qy 1618 AGCCTGAAAGCCTTGGAGAGGCGAGTGGTATCCTAGAGATGCTGCTCTTCCAT 1677  
 Db 1646 AGCCTGAAAGCCTTGGAGAGGCGAGTGGTATCCTAGAGATGCTGCTCTTCCAT 1705  
 Qy 1678 AATATATTTATTAACAACCTTTATATGAAACATCAGTCTTCACTGAGAGATGTAT 1737  
 Db 1706 AATATATTTATTAACAACCTTTATATGAAACATCAGTCTTCACTGAGAGATGTAT 1765  
 Qy 1738 GCAGTGGCAAAATTAAGTGTGATTCATGATGCAATCTTGAATGCCAATGATATGAC 1797

Db 1766 GCAGTGGCAAAATTAAGTGTGATTCATGATGCAATCTTGAATGCCACATGATATGAC 1825  
 Qy 1798 ACCCAAGTAGGGAACGAGGACTCAAGCTTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAT 1857  
 Db 1826 ACCCAAGTAGGGAACGAGGACTCAAGCTTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAT 1885  
 Qy 1858 GCAAGAGCCATTTTGAAGAGACCCCGCAGTCATACCTATGATGAAGTACTTCAATGTTA 1917  
 Db 1886 GCAAGAGCCATTTTGAAGAGACCCCGCAGTCATACCTATGATGAAGTACTTCAATGTTA 1945  
 Qy 1918 GATTGATTACTGAAGAGATATCTTGTGTCATGAAGAGATGTCATGAAGAGATGTCATGAAG 1977  
 Db 1946 GATTGATTACTGAAGAGATATCTTGTGTCATGAAGAGATGTCATGAAGAGATGTCATGAAG 2005  
 Qy 1978 TCTATTTTCACTGACACAGATGTCACAGTGTGATGATGATGATGATGATGATGATGATG 2037  
 Db 2006 TCTATTTTCACTGACACAGATGTCACAGTGTGATGATGATGATGATGATGATGATGATG 2065  
 Qy 2038 GATCAGGGTAAAGTAAAGCGTGGTACCCACATGTTGTTGCTTAACCTCATAGT 2097  
 Db 2066 GATCAGGGTAAAGTAAAGCGTGGTACCCACATGTTGTTGCTTAACCTCATAGT 2125  
 Qy 2098 ATCTATTCAGAAATGTGGCATACACAGAGACCGGTGTGAGAAACATGATTAACCCCAA 2157  
 Db 2126 ATCTATTCAGAAATGTGGCATACACAGAGACCGGTGTGAGAAACATGATTAACCCCAA 2185  
 Qy 2158 TGGGAAGC 2165  
 Db 2186 TGGGAAGC 2193

RESULT 5  
 ID ABL01991  
 ID ABL01991 standard; cDNA: 2182 BP.  
 XX ABL01991;  
 AC 26-MAR-2002 (first entry)  
 DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 455.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX OS Drosophila melanogaster.  
 PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US09231.  
 XX PR 23-MAR-2000; 2000US-191637P.  
 XX PR 11-JUL-2000; 2000US-0614150.  
 XX PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PMD, Myers BW;  
 XX WPI; 2001-656860/75.  
 DR P-PSDB; ABB57888.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX Claim 1; SEQ ID NO 455; 21pp + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB16175), expressed DNA  
 CC sequences (AB16176-AB16175) and the encoded proteins  
 CC (AB57737-AB57737).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIRO  
 CC at ftp.wipo.int/pub/published\_pcf\_sequences.

XX  
 SQ Sequence 2182 BP; 491 A; 588 C; 633 G; 470 T; 0 other;

Query Match 24.1%; Score 564.8; DB 23; Length 2182;  
 Best Local Similarity 58.5%; Pred. No. 5.9e-152;  
 Matches 1011; Conservative 0; Mismatches 702; Indels 15; Gaps 1;

```

Qy 354 GAAATCATTAAGCAATCTTTATGTCGCGCCAAAGACAGGCCAGATCTACGAGC 413
Db 358 GGAATGCTGCGCGCATATGAGCCATACATCTGCGAAAGAGATCCACTGGTGGAAA 417
Qy 414 TAGAGTTGCCATTTCGTGGGATTTTGGGTGTGCAAGGCCATGAATATTGTGGTCC 473
Db 418 GCGGGTGGGATTTTCCTGGGTCTGTGGCTGCTCAAGCTGCTGACCGCTGTGTCGC 477
Qy 474 CTTCAATGTTTAATATCTGTAGACAGCTCAACAGATGTGGGAAACATGCTAACT 533
Db 478 CTTCTGTTCAAGAGAGCTGTGACAC-----CGATGACACGCTGAACAT 522
Qy 534 GAGTATGACCAAAATACAGTTGCAACATGCGCAACAGCAGTTGATTTGCTATGCTGT 593
Db 523 GGAACATGCGCGCGAGTGGGTCTCTCCGCAACACCGCAGATGTTGGATATGAGCAT 582
Qy 594 ATCAAGAGCTGAGCTGCTTTTATACGAGATTGAAAATGCAATATTGGCAAGGTAC 653
Db 583 TGCTAGAGCAATGGCGCGGCTTTATGATGGCGAATGCAAGTGTGGCCAAAGTGGC 642
Qy 654 CCAGAAATCAATCCGAAGATAGCCAAAGTCTTTCCATCTTCAACCTGGATCT 713
Db 643 CCACCACTGATCCGAAGATGCGCAGCATGTCTCTGCACTGACACACCTGGATCT 702
Qy 714 GGGTTTCACTGAGCAGACAGACGCGAGCTTTATCTAAGGCTATTGACAGAAACAG 773
Db 703 GGCCTTCCACCTGAAACAAACAACTGAGCACTGTCAAAAGACATGATCGAGATCGAG 762
Qy 774 GGGTATCAGTTTGTCTGAGTCTTGGTATTTAATCTTCTCCATCATGTTGAACT 833
Db 763 GGGCATTTAACTTTGTCTCTCCGCAATGGTCTTCAACATTTGCCACATCTTTGAGCT 822
Qy 834 GATGCTTGCAATGCTGTTTGTATTTAACAATGCGGTGCCAGTTGCTTGGTAACCT 893
Db 823 GGCCTCTGTCAGATATCTGGAGTGAAGTGTGGCTGGCTGTGCTGTGACAT 882
Qy 894 TGGAACTTTGGTACATACAGACATTCACAGTTGCACTCACAGGAGAGAACTAGATT 953
Db 883 GGGCTGCGTTGGCATATAGCTGCTGACACTGTGAGTGTGACCCAGTGGCAGCCAGTT 942
Qy 954 TAGAATAGAATAGAACAAAGAGATATGATGATGATGCTGCTTATGACTCACTGCT 1013
Db 943 CGGTGATTTATAGAACAGCGGAGAAAGAGCCGCGCAACAGGCGTAGACTCCGCTGAT 1002
Qy 1014 GAATATGAATCTGTAAGTATTTTAAATGAAGAATGAAGACAGAGATATGATGAG 1073
Db 1003 CAACACGAGAGCGGTGAATATCTTCAACAGAAAGTACGAGCGGCTGTGACACAGA 1062
Qy 1074 ATTTTGAAGAGATGAGATGCTTCAATTTGAAAAGTACCTTCACTGGCTATGCTGAA 1133
Db 1063 GGTGCTGAAGAGATGAGCGCGCGCTTGAAGACTACAGTCTGCTGTCTGCTCA 1122
Qy 1134 CTTTGGTCAAGTGTCTATTATGATGTGGTTTACAGCTTATATGCTGTGCTGCACTCA 1193
Db 1123 CTTTGGCCGAGAGCGCAATTTTCAAGCGCAGCTGATGATTAATGATGATTTGGCCGCA 1182
Qy 1194 GGGAAATGTGAGAGTACCTTACTGTGAGATCTAGAAATGAGTGAATGAGCTGCTTTT 1253
Db 1183 GGAAGATTGCCAAGGCAACATGACGTTGGAGATTGTGATGTCACAGCCCTGCTCTT 1242

```

```

Qy 1254 TCAGCTTTCATTAACCCCTGAACTTTCTGGAACTGTATATAGAGACTAGACAGACT 1313
Db 1243 CCAGCTTCGATCCCCCTCGGCTTTCTGGGAGAGTATATCTGATGAGTGGACAGGCTTT 1302
Qy 1314 CATGATATGAAACCTTGTTTACTTACTCAAGTAGACCCCAATTAAGCAAGT 1373
Db 1303 GCTGACATGCGGCGCATGTACGTTGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1352
Qy 1374 GATGCACTCTCCCTTCAGATCACACACAGACGTAACGCTGGCTTTGATTAATGCA 1433
Db 1363 CAATGCCAGGCCCTGTGTCGTGACACACACTAATCATCTTGAATTCGCAAGTGAAG 1422
Qy 1434 TTTGATATCAATGAGCGCAAGAAAGTCTTATGTAATATCTTTGAATGCTGCAAG 1493
Db 1423 CTTGAGATGCAAACTTGGCAAGCCATTTTCCGGAGCTCAGTTTCAACATCCGCGCG 1482
Qy 1494 AAAAGAGTGGCCATTTGATGAGTATGAGTGTGAGGAGAAAGCAATATGAGGCTATT 1553
Db 1483 TMAAAGCTGCGCATTTGTGGGCGGCTCTGCTCAGGAAATCTGTATGTTGCGCTTT 1542
Qy 1554 ATTTGCTTATGAGCTTCAAAAGGATGATTTATCTTGTGTCGTCMAAATATACAGA 1613
Db 1543 CTTGCGCTTCTTTAGACCAAACTGTGTAAGTGTGATGCTGTCGCAAGATATCAAGCG 1602
Qy 1614 TGTAGCCTGGAAGCCTTGGAGGCGAGTGGAGTGTGATCTCAGATGCTGTCTCTT 1673
Db 1603 CGTGGATTTGAGAGCCTGCGCAAGTTATGCAAGTGTGCGCAGGATTCGCTGATATT 1662
Qy 1674 CCATATATCTATTTATTAACAACCTTATATGAAACATCAGTCTTCACTGAGGAAGT 1733
Db 1663 CCACACACCATGACAGCAATATATCACTATGGAACCTGTCAAGTCCACGCAAGAT 1722
Qy 1734 GTATGAGTGGCAAAATTTAGCTGACTTATGATGCAATTTCTTGAATGCCACATGATA 1793
Db 1723 GCAAGATCTCCCGCATAGCTGTATTTGATGATGATGATGATGATGATGATGATGAT 1782
Qy 1794 TGAACCCAGTAGGAGAGAGAGAGATCTCAGCTTTCAAGAGAGAAAGCAAGATAGC 1853
Db 1783 CTCACATGAGTAGGAGAGAGAGAGATTTAAAGTATCCGCTGCGCAGAGAGAGAG 1842
Qy 1854 AATTGCAAGACCATTTTGAAGAGACCCCGCATATCTGATGATGATGATGATGATC 1913
Db 1843 CATGCTAGGCGCATATCTGAAAGAACCCCATCTGATCTTTGACAGGCTTACAGCAG 1902
Qy 1914 GTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1973
Db 1903 TGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1962
Qy 1974 AACTTCTATTTTCAATGACACAGATTTGCAACAGTGTGATGATGATGATGATGAT 2033
Db 1963 CACTAGCATTTTGAATTTGCGCCACGCTCTCTGACAGTCAAGATGCGGACGAGATTCTGT 2022
Qy 2034 CTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2081
Db 2023 GCTCGAAGACGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2070

```

RESULT 6  
 AB101989  
 ID AB101989 standard; cDNA; 2539 BP.

XX AB101989;  
 XX 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 449.  
 DE Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX Drosophila melanogaster.  
 OS  
 XX  
 PN W0200171042-A2.

XX 27-SEP-2001.  
 PD  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li FWD, Myers EW,  
 XX  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB57886.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 449; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB160511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB857737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 2539 BP; 580 A; 677 C; 708 G; 574 T; 0 other;

Query Match 24.1%; Score 564.8; DB 23; Length 2539;  
 Best Local Similarity 58.5%; Pred. No. 6,4e-152;  
 Matches 1011; Conservative 0; Mismatches 702; Indels 15; Gaps 1;

QY 354 GAAATCATAAAGCAATGCTTCTATATGTGGCCCAAGAGCCAGATCTACGAGC 413  
 DB 715 GGAATGCTGGTGCATGATGCTTACATCTGGCCAAAGAGATCCACTGGCGAAA 774  
 QY 414 TAGAGTGGCCATTTGCTGGGATTTTGGGTGTGCAAGGCCATGAATATGTTGCTTC 473  
 DB 775 GCGGGTGGGATTTTCCCTGGGTCTGTGGTGTGCTCCAGCTGTGACCGCTGTGCTCC 834  
 QY 474 CTTGATGTTAAATATGCTGTGAACAGCTCAACCAATGTGGGAAACATGTAACCT 533  
 DB 835 CTTCTGTTCAAGAGAGCTGTGACA-----CGATGACAAAGCTGAACAT 879  
 QY 534 GAGTATGACACCAATACAGTTGCAACATGCGAAGCAAGCTTGTGATGCTATGATGT 593  
 DB 880 GGAACACTGCCCGGATGGGCTCTCTCCGACGACCGACTGATGTTGGATATGGCAT 939  
 QY 594 ATCAAGAGCTGAGCTGCTTTTAAAGAAATGCAATGCAATGTTGGCAAGTAGC 653  
 DB 940 TGCATAGAGCAAGTGGCGGCTTAATGAGTTGCGGAATGCAATGTTGGCAAGTAGC 999  
 QY 654 CCAGATTCATCCGAAAGATAGCCAAAATGCTTTTCTCATCTTCAACCTGATCT 713  
 DB 1000 CCACCACTGATCCGAAAGATCGCACCAATGTTCTGATCTGACCAACCTGATCT 1059  
 QY 714 GGGTTTCACTGAGCAGACGAGCGGAGCTTATCTAAGGCTTACAGAGGAAAG 773  
 DB 1060 GGGCTTCACTGAGCAGACGAGCGGAGCTTATCTAAGGCTTACAGAGGAAAG 1119  
 QY 774 GGGTATCAGTTTGTCTGAGGCTTGTGATTAATCTTCTTCCATCAATGTTGAAGT 833  
 DB 1120 GGGGATTAACCTTGTGTCTCCGCAATGTTTCAACATGTTGCTCCATCTTTGAGCT 1179  
 QY 834 GATCTGTCAATGTTGTTTGTATTAACAATGCGGTGCCAGTTGCTTTGTAACCT 893

DB 1180 GGCCCTGTGTCCAGTATCTCTGGAGTGAAGTGTGGCTTGCCCTGCTGTGTACAGAT 1239  
 QY 894 TGAACACTTGTGATACATACAGACATTCACAGTTGCAGTCAACGCTGAGAACTAGATT 953  
 DB 1240 GGGCTGTGCTGGCATATAGCTGTGCTACCTGAGAGTGTACCAAGTGGCGACCCAGTT 1299  
 QY 954 TAGAATGAAATGAACAAGCAAGTAATATGATGCAAGTATGCTATTAAGCTCACTGCT 1013  
 DB 1300 CCGTGTATTTATGAAACGAGCGGAGAAACGAGCGCGGCAACAAGCCGTAGACTGTGAT 1359  
 QY 1014 GAATTAAGAACTGTGAAGTATTTTAATATGAAGTATGAAGCAACAGATATGATGG 1073  
 DB 1360 CAATTAAGAAACGCTGAGTAACTTCAACAACGAAAGTACAGCGGGTGTCAACAAGA 1419  
 QY 1074 ATTTTGAAGACGATATAGACTGCTTCAATGAAGAAGTACTTACTGTGCTATGCGAA 1133  
 DB 1420 GGTCTGAAGAAATATAGAGCGGCGAGCTTGAAGACTAGTCCAGTTCGCTCTGCTCAA 1479  
 QY 1134 CTTTGTCAAAAGTCTAATTTTCAAGTGTGCTTTAACAAGTAAATGCTGTCCAGTCA 1193  
 DB 1480 CTTTGGCCAGAACGCGCATTTTACAGACGCGACTGAGCTGATATGATTTGGCCGCAA 1539  
 QY 1194 GGAATATGTGCGAGTACCTTACTGTGAGATCTAGTAAATGATGATGACTGCTTTT 1253  
 DB 1540 GGAATATGCGCAAGGCAACATGACGTTTGAATGATGATGATGATGATGATGATGAT 1599  
 QY 1254 TCAGCTTTCATTAACCCCTGAACCTTTCGGGAACTGTATATAGAGACTAGCAAGCACT 1313  
 DB 1600 CAGCTCTCATATCCCTCGCTGCTTGTGCGAGTGTATTCGAGAGTGGACAGGCTTT 1659  
 QY 1314 CATAGATATGAACACTTGTGTTAATCTTACTCAAGATAGACACCAATTAAGAANAAGT 1373  
 DB 1660 GGTGATCATCGGGCCATGTTTCAAGTATGATGATGATGATGATGATGATGATGATGAT 1719  
 QY 1374 GATGCAATCTCCCTTCAATTCACACAGACAGTACCTGCTGCTTATATATGCA 1433  
 DB 1720 CAATGCGCAACCCCTGTGTGTGACACACTAATCATCATTAAGTATCCGCAACGAG 1779  
 QY 1434 TTTTGAATACATTGAGGGGCGAGAAAGCTTATGGAATATCCTTGAATCCCTGAGG 1493  
 DB 1780 CTTGAGATGAACCTGAGCAAGCCATTTTCCGGACCTGATTTCAACATACCCGCGG 1839  
 QY 1494 AAGAAAGTGGCAATGTATGAGATGAGTGGTTCAGGAAAAAGCAATATGAGGCTATT 1553  
 DB 1840 TAAAGAGTGGCAATGTATGAGATGAGTGGTTCAGGAAAAAGTGTGATGAGGCTTCT 1899  
 QY 1554 ATTTGCTTATAGAGCTTCAAAAGGTAGCATTTATCTTGTGCTCAAAATATACAGA 1613  
 DB 1900 CTTCCGCTTCTTGAAGCAAACTGTGTAAAGTGTGATGAGTGGCGCAGATATACAGCG 1959  
 QY 1614 TGTAGCTGGAAAGCTTCCGGAGGCGAGTGGAGTGTACCTCAGAGTCTGCTCTT 1673  
 DB 1960 GGTGATTTGGAAGCTGTGCAAGGTATATGCAAGTGTGCGCAGAGATTCGGTCTATT 2019  
 QY 1674 CCATTAATCTATTATTAACAACCTCTTAATATGAAACATCAGTCTTCAACCTGAGAA 1733  
 DB 2020 CCACAAACCATGAGACAAATATCACTATGSCAACTGTCCAAATCCACGCAAGAGT 2079  
 QY 1734 GTATGCAATGCGCAAAATATAGCTGACTTATATGATGCAATCTTGAATGCAATGATA 1793  
 DB 2080 GCAGAAATGCTGCGCGGATGCTGATTTGATGATGATGATGATGATGATGATGATGAT 2139  
 QY 1794 TGACACCAAGTATGAGGAGGAGAGGAGCTTCAAGCTTTCAGGAGGAGAAAGCAAGAGT 1853  
 DB 2140 CTCACCTCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2199  
 QY 1854 AATTGCAAGGCAATTTTGAAGAGCCCGCAGTCACTATCTATATGAAGCTACTTATC 1913  
 DB 2200 CATGCGTATGAGGCACTATGAGAGACCCCAATTCATCTTTTGAAGAGGCTCAAGAGAG 2259  
 QY 1914 GTTAAATGATTAATGAGAGACTATTTCTGCTGCAATGAGAGATGTGCTCAACAGAG 1973  
 DB 2260 TCTGAGCTGATCAGGAGCATTAATCTGAGGCTGAGCCGCGCTAATCTCAGAGAGG 2319



Oy 1974 AACTCTATTTCATTCGACAGATTGTCACAGTGGTTGATGAGATGAATCATTTGT 2033  
Db 2320 CACTGACATTTTCATTCGACAGATTGTCACAGTGGTTGATGAGATGAATCATTTGT 2379  
Oy 2034 CTTCGATCAGGAGTGAAGTGAAGCCGAACTGTGTACCCACCATGTTTGT 2081  
Db 2380 GCTCGAAGACGAGCGTGTAGTGAACGTGGACCCATTCGAGCTGTCT 2427

RESULT 7  
AAH05966  
ID AAH05966 standard; cDNA; 620 BP.  
XX  
XX AAH05966;  
AC  
XX 26-JUN-2001 (first entry)  
DT  
XX  
XX Human cDNA clone (5'-primer) SEQ ID NO:2801.  
DE  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
KM  
XX Homo sapiens.  
OS  
XX EP1074617-A2.  
PN  
XX 07-FEB-2001.  
PD  
XX 28-JUN-2000; 2000EP-0116126.  
PF  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
PA  
XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
DR  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 1; SEQ ID 2801; 2537bp + CD ROM; English.  
PS  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification, where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination  
CC of the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH1363 to AAH1368 and  
CC AAH1363 to AAH1874 represent human cDNA sequences; AAH82446 to  
CC AAH93893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
XX

SQ Sequence 620 BP; 158 A; 149 C; 156 G; 147 T; 10 other;  
Query Match 23.5%; Score 551.6; DB 22; Length 620;  
Best Local Similarity 97.0%; Pred. No. 1.8e-148;  
Matches 578; Conservative 0; Mismatches 14; Indels 4; Gaps 2;  
Oy 1 ATGGGCGTCTCGGATGATTCATTTGGCGCTGGGCGGCGGCGGCTTTTGGAAAG 60  
Db 26 ATGGCGCTGCTCGGATGATTCATTTGGCGCTGGGCGGCGGCGGCTTTTGGAAAG 85  
Oy 61 CGCCGGCATCTCCGAGATTTGATCCGGCTTTAGTCTCTGTACCGGCTCAGATCCGAG 120  
Db 86 CGCCGGCATCTCCGAGATTTGATCCGGCTTTAGTCTCTGTACCGGCTCAGATCCGAG 145  
Oy 121 TGGAGGCGCATCAACTCGGCGGCGCTTGGGAAACCGCTGAGCGTAC---CAGATTCCAGAG 177  
Db 146 TGGAGGCGCATCAACTCGGCGGCGCTTGGGAAACCGCTGAGCGTACCGAGATTCAGAN 205  
Oy 178 TCATTTAAAAGTATCAGATGCGAGAGATTGGGAAAAGCAATTGAGACAGATTCTTAGAT 237  
Db 206 TCATTTAAAAGTATCAGATGCGAGAGATTGGGAAAAGCAATTGAGACAGATTCTTAGAT 265  
Oy 238 GCTGCAAAAGGCTCTCCAGATATGCGCCTGACTGATAGAAAAGAGACATGTTGCAATGCTCAT 297  
Db 266 GCTGCAAAAGGCTCTCCAGATATGCGCCTGACTGATAGAAAAGAGACATGTTGCAATGCTCAT 325  
Oy 298 GCAGAGAGAGAGCTCCACACAGACCCAAAGAGGTTTAAAGATGTTAGTCTCGGAAA 357  
Db 326 GCAGAGAGAGAGCTCCACACAGACCCAAAGAGGTTTAAAGATGTTAGTCTCGGAAA 385  
Oy 358 ATCATTTAAAAGCAATGCTTTCTTATGTTGTCGCCCAAGACGCGCAGATCTACGAGCTAGA 417  
Db 386 ATCATTTAAAAGCAATGCTTTCTTATGTTGTCGCCCAAGACGCGCAGATCTACGAGCTAGA 445  
Oy 418 GTTGCCATTTGCTGGGATTTTGGGTGTGCAAAAGGCATGAATATTGTTGCTTCCCTTC 477  
Db 446 NTGCCAATTTGCTGGGATTTTGGGTGTGCAAAAGGCATGAATATTGTTGCTTCCCTTC 505  
Oy 478 ATGTTTAAATATGCTGTAGAGACGCTCAACGATGTCGGAAACATGCTGAACCTAGT 537  
Db 506 ATGTTTAAATATGCTGTAGAGACGCTCAACGATGTCGGAAACATGCTGAACCTAGT 565  
Oy 538 GATGACCAATAATAGTTGCAACCATGSCAACAGACAGTTGCTGATGCTATGCTGT 593  
Db 566 GATGACCAATAATAGTTGCAACCATGSCAACAGACAGTTGCTGATGCTATGCTGT 620

RESULT 8  
AAH34971  
ID AAH34971 standard; cDNA; 807 BP.  
XX  
XX AAH34971;  
AC  
XX 03-SEP-2001 (first entry)  
DT  
XX  
XX Human colon cancer antigen encoding cDNA SEQ ID NO:2053.  
DE  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KM colorectal carcinoma; ss.  
KW  
XX Homo sapiens.  
OS  
XX W0200122920-A2.  
PN  
XX  
XX 05-APR-2001.  
PD  
XX  
XX 28-SEP-2000; 2000MO-US26524.  
PF  
XX 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX





```

Db 4102 TTGAGCTGGCCCTGCTGTCAGTATCTGGGAGTGAAGTGGCCCTGCGCTTCGCTGCTG 4161
Oy 887 TAACCTTGGAAACATTGGTATCATACAGCATTCAGATTGCTCACTCACGCGTGGAGAA 946
Db 4162 TCGAGATGGGCGTGGCTGGCATATAGCTGCTTACACTGTGATGTGACCCAGTGGCGGA 4221
Oy 947 CTAGATTTAGAAATAGAAATGAAACAAGCATTAATGATGACAGTAAATGCTGTATAGACT 1006
Db 4222 CCCAGCTCCGTATTTATTAAGAACCGAGCGGAGAACGAGGCCGCGCAACAAGCCGTAGACT 4281
Oy 1007 CACTGCTGAATATATAAAGTGTGAAGTATTTTAAATTAAGAAATATGAAGACAGACAT 1066
Db 4282 CCGTATCAACTACAGAGAGGTGAAGTACTTAACAACAAGAAAGTACGAGGGGCTGTGCT 4341
Oy 1067 ATGATGATTTTGAAGAGTATGAGACTGCTTCAATTGAAAAGTACCTCTACTCTGGCTA 1126
Db 4342 ACAAGAGGTGCTGAAGAGTATGAGGCGGCCAGCTTGAAGACTAGCTCCTCACTTCGCTC 4401
Oy 1127 TGTGAACCTTTGGTCAAAAGTGTATTTTCAAGTGTGCTTAAACAGCTATATATGATGCTG 1186
Db 4402 TGTCAACTTTGGCCAGAACGCCATTTTCAGACGCGCACTGAGTCTGATTAATGATTTGG 4461
Oy 1187 CCAAGTCAAGGAATTTGGGAGAGTACCTTACTGTTGGAGATCTAGTAATGTGAATGAC 1246
Db 4462 CCGCAAGAGAGATGCCCAAGGCAACATGACGCTTGAAGATTTGGATGTGCAACGCC 4521
Oy 1247 TGTCTTTTCAAGCTTTCATTAACCTTGAACCTTCTGGGAACTGTATATAGAGACTAGAC 1306
Db 4522 TGTCTTTCAGCTTCTGATATCCCTTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 4581
Oy 1307 AAGCACTCATAGATATGAACACCTTGTCTTACTCTACTCAAGGTAGACACCCAAATTAAAG 1366
Db 4582 AGGCTTTTGTGACATGCGGCGCCATGTTCACTGTATGAACGTGACAGTACATTCAGA 4641
Oy 1367 ACAAGTGAATGGCATCTCCCTTCAGATACACACAGACACTACCGTGGCTTTGATA 1426
Db 4642 CAGCGCCCAATGCCAGCCCTGCTGTGTGACACACTAATCATCTATGATGAGTTCGCA 4701
Oy 1427 ATGTGATTTTGAATCATTTAGAGGCGCAGAAAGTCTTGTAGAAATATCTTTGAAGTCC 1486
Db 4702 AGGTGATGCTTCAGATACGAACTGGCAGAGCCATTTCCGGAGCTCACTTTCACCATAC 4761
Oy 1487 CTGCAAGAAAGAAAGTGGCATTGTAGAGAGTGTGGGTCAAGGAAAGACACAATATGTA 1546
Db 4762 CCGCGGTAATAAAGTCGCGCATTTGTGGCGGCTCTGGCTCAAGAAATGCTGATGTGC 4821
Oy 1547 GCGTATATTTTGGCTTCTATAGAGCTCAAAAAGGTATGCTTATCTGCTGTCAAAAAT 1606
Db 4822 GCGTCTCTTCGCTTCTTGTGAGCCAAACTCTGTAAAGTGTGATCGGTGGCCAGGATA 4881
Oy 1607 TACAAGATGTAGAGCTGGAAGGCTTCGGAAGGCAAGTGGAGTGGTACTCAGATGCTG 1666
Db 4882 TCAAGCGCGGTGATTTGGAGAGCTGCGCAAGTTATTCAGATGTGTGCGCGCAGATTCG 4941
Oy 1667 TCCCTCTTCATATACTATTTATACACCTCTTATATGAAAACATCAGTGTCTTCACTG 1726
Db 4942 TCTATTCACAACACATCGAGACAAATATCATATGGAACCTGTCCAGTCCCAAG 5001
Oy 1727 AGGAAGTATGAGAGTGGCAAAATAGTGTGACCTTATATATGATGATCTTCTGAAATGCC 1786
Db 5002 CAGAAGTGAAGATGTGCGCCGATGGCTGATTTGATATATCCATCAATGAACTGGCCAG 5061
Oy 1787 ATGATATATGACCCCAAGTATGGGGAACGAGACTCAAGCTTTCAGAGGAGAAAGCAAA 1846
Db 5062 GACAGTACTCTCAGTGAAGGAGGAGAGTGAATTAAGTTATCCGTTGCGAGAGAGAC 5121
Oy 1847 GAGTATGATTCAGACCAAGTATGGGGAACGAGACTCAAGCTTTCAGAGGAGAAAGCTTA 1906
Db 5122 GCGTTGCCATCGCTAGGCGCATATCTGAAGAACACCCCAATTCATCTTGAAGAGGCTTA 5181
Oy 1907 CTTCATCGTAAATTCGATTAATCTGGAAGAGACTATTTCTTG 1945
Db 5182 CAAGCAGTCTGAGCTCGATCACCAGACATGTAGTGTG 5220

```

```

RESULT 10
ABL01990
ID ABL01990 standard; cDNA; 6541 BP.
XX
XX ABL01990;
XX
XX
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 452.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW,
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB57887.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 452; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 6541 BP; 1635 A; 1580 C; 1654 G; 1672 T; 0 other;
XX
Query Match 20.5%; Score 480.6; DB 23; Length 6541;
Best Local Similarity 59.6%; Pred. No. 2.2e-127;
Matches 810; Conservative 0; Mismatches 549; Indels 0; Gaps 0;
Oy 587 ATGTGATATCAAGAGCTGAGCTGCTTTTAAAGAGTTCGAATGACGATTTGGCA 646
Db 3869 ATGGCATTGCTTAGAGCCAGTGGGGGCGCTTAAATGATTCGGAAATGCAATGTTGGCA 3928
Oy 647 AGGTAGCCCAAGATTCATCCGAAGATAGCCAAATATGCTTTCTCCATCTTCAACAC 706
Db 3929 AGGTGGCCCAACATCCGATCCGAAGATGCGCAGCAATGTGTTCTGCATCTGCAACAC 3988
Oy 707 TGGATCTGGGTTTCACTGACGACAGACGAGAGCTTTATCTAAGCTATTTGACAGAG 766
Db 3989 TGGATCTGGGTTTCACTGACGACAGACGAGAGCTTTATCTAAGCTATTTGACAGAG 4048
Oy 767 GACCAAGGAGTATCAGTTTGTCCGAGTGTGTTGATTAATCTTCCCATCATGT 826
Db 4049 GATCAGGAGGATTAATCTTGTGCTTCGCGCATGTGTTCAACATTTGGCCACTATCT 4108

```

Qy 827 TTGAAGTACCTGTCAGTGTGTTTGTATTAACAAATGCGGTCCAGTTGCTTTGG 886  
 Db 4109 TTGAGCTGGCCCTGTCAGTATCCCTGGAGTGAAGTGGCCCTTCCTGCTGTG 4168  
 Qy 887 TAACCTTGGAACTTGGTACATACAGAGATTACAGTTGACATGACAGGTGGAGAA 946  
 Db 4169 TCAGCATGGGCTGGCTTGGCATATACGCTGCTTACCTCTGAGTGTGACCCAGGGGCA 4228  
 Qy 947 CTAGATTAGAAATGAATGAACAAAGCATTAATGATGACGTAAATGCTTAAACT 1006  
 Db 4229 CCCAGTTCGGTATTTATTAAGAACGAGGCGAGAAAGAGGCGGCAACAGGCGTAGACT 4288  
 Qy 1007 CACTGCTGAATTGAAACTGTGAATTTTAAATGAAGAATGAACACAGAT 1066  
 Db 4289 CGCTGATCACTACGAGCGGTGAAGTACTTCAACAGAAATAGAGCGGGTGGCT 4348  
 Qy 1067 ATGATGATTTTGAAGCGTATGAGACTGCTTATGAAAGAAAGTAACTTACTGCTTA 1126  
 Db 4349 ACAACGAGGTGCTAAGAAATGAGGGGCGACCTTGAAGACTAGCTCCAGCTCCCTC 4408  
 Qy 1127 TGCTGAACCTTGTCAAAAGTGTCTATTTTTCAGTGTGCTTAAAGTAAATGCTGCTG 1186  
 Db 4409 TGCTCAACTTTGGCGAAGACCCATTTTCAGCAGCGCAGCTGATCTGATTAATGGTATGG 4468  
 Qy 1187 CCACTCAGGGAATTTGGCAGGTAACCTTACTGTTGAAGATCTAGTAATGCTGAATGAC 1246  
 Db 4469 CCGCAAGGATTTGCCCAAGGCAACATGACGTTGAGATTTGGTATGCTCAACGCC 4528  
 Qy 1247 TGCTTTTACGCTTCAATTCCTGAACTTTGCGGAACCTTATAGAGACTGAC 1306  
 Db 4529 TGCTTTCAGCTCTGATCCCTCGGTTTCTGGGAGGTTTATGAGAGGTGGAC 4588  
 Qy 1307 AAGCACTCATGATATGAACACTTGTCTACTCAAGGTGAGACCCCAATTAAG 1366  
 Db 4589 AGGCTTTGCTGAGATGGGGCCATGTTCACTGATGAAGTGAAGTGAAGTGAAGTGAAG 4648  
 Qy 1367 ACAAGTGAATGCAATCTCCCTTCAATCACACAGACAGTACCGTGGCTTTGATA 1426  
 Db 4649 CAGCGGCAATGCCAGCCCTGTTCTGAGACACACTAATCTCATGATGATGGTCCGA 4708  
 Qy 1427 ATGTCATTTTGAATATGAGGGCCAGAAAGTCTTATGGAATTCCTTTGAATCC 1486  
 Db 4709 ACGTACCTTGAATGAGAACCTGCAAGCCCATTTTCGGGACCTCAATTTCCACATAC 4768  
 Qy 1487 CTGCAAGAAAGAAAGTGGCCATTTAGAGGTAGTGGTCCAGGAAAGCAATATGTA 1546  
 Db 4769 CCGCGGTAAGAAAGTGGCCATTTGAGGCGGCTGTGCTCAGGAAATTCCTCATGATGTC 4828  
 Qy 1547 GGCATTAATTTGCTTATGAGCCTCAAAAGGTAGCATTTATCTGCTGTCAAAATA 1606  
 Db 4829 GCTTCTCTTCGCTTCTTTGAGCCAAACTCTGTTAAAGTGTGATGGTGGCCAGATA 4888  
 Qy 1607 TACAAGATGAGCCTGGAAGCCTTGGAGGCGAGTGGAGTGTACTTCAGATGCTG 1666  
 Db 4889 TCAGCGCGGTGATTTGAGAGCCTGGCAAGGTTATTTGAGTGTGTCGAGGATTCGG 4948  
 Qy 1667 TCCCTTCATTAATTAATTAATTAACAACTTATATGAAACATGATGCTTCACTG 1726  
 Db 4949 TGCATTTCCACACATGAGACAAATTCATTTGCACTTGGCACTTCCAGATGCCACG 5008  
 Qy 1727 AGGAAGTATGAGTGGCAAAATTAAGTCTGACTTCAATGATGCAATTTCTGAAATG 1786  
 Db 5009 CAGAAGTGAAGATGCTGCCGCGCATGGCTGATTTGATGATCATCATGATGCTGGCAG 5068  
 Qy 1787 ATGATATGACACCCAGTGAAGGAGAGAGATCAAGCTTTACGAGAGAAAGCAAA 1846  
 Db 5069 GACAGTACTCCACTCAGATGAGGCAAGAGTGAATTAATCGGTGGGAGAGAGAGC 5128  
 Qy 1847 GAGTGAATTTGCAAGGCAATTTTGAAGAGACCCCGAGTCAATCTCTATGATGAGCTA 1906  
 Db 5129 GCGTGGCATTCGTAGGGCCATCTGAAGAAACACCCCATTTCTGATCTTTGACAGAGCTA 5188  
 Qy 1907 CTTCATGTTAGATTGATTAATGAGAGACTATTTTG 1945

Db 5189 CAAGAGTCTGAGCTCGATCAACGAGCATGATGATGTTG 5227  
 RESULT 11  
 AAA45541  
 ID AAA45541 standard; cDNA, 437 BP.  
 XX  
 AC AAA45541;  
 AC  
 DT 21-AUG-2000 (first entry)  
 XX  
 DE Human secreted expressed sequence tag SEQ ID NO:2116.  
 XX  
 XX Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;  
 KM expressed sequence tag; SEST; probe; chemotactic; proliferative;  
 KM immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;  
 KM thrombolytic; antiinflammatory; cyostatic; antibacterial; antifungal;  
 KM antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;  
 KM anticancer; osteopathic; neuroprotective; nootropic; antiparalytic;  
 KM cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
 KM vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
 KM insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
 KM lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
 KM central nervous system disorder; Alzheimer's disease; stroke;  
 KM Parkinson's disease; Huntington's disease; coagulation disorder;  
 KM haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
 KM tumour; infection; depression; psoriasis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20021991-A1.  
 XX  
 PD 20-APR-2000.  
 XX  
 XX 15-OCT-1999; 99WO-US24206.  
 PF  
 XX 15-OCT-1998; 98US-010436.  
 PR  
 XX  
 PA (GENEY ) GENETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Bowman MR;  
 XX  
 DR WPI; 2000-317938/27.  
 PT Isolated polynucleotides, and encoded proteins, comprising secreted  
 PT expressed sequence tags (SESTs), useful for treating various disorders  
 PT such as autoimmune, infectious, and central nervous system disorders -  
 XX  
 PS Claim 1; Page 731; 803pp; English.  
 XX  
 CC AAA43426 to AAA45925 represent specifically claimed secreted expressed  
 CC sequence tags (SESTs), isolated from human, mouse, chicken and rat  
 CC tissue sources. The SESTs can have a range of activities depending on  
 CC the tissues they were isolated from. The activities include:  
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;  
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;  
 CC antiasthmatic; vulnery; anticancer; osteopathic; neuroprotective;  
 CC nootropic; antiparkinsonian; antiparalytic; cerebroprotective;  
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene  
 CC therapy and in vaccines. The SESTs are useful as probes for the  
 CC identification and isolation of full-length cDNAs and genomic DNA  
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs  
 CC are useful in assays for determining biological activity and raising  
 CC antibodies. They may be useful for treatment of autoimmune disorders  
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
 CC osteoporosis, osteoarthritis, central nervous system disorders  
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
 CC disease), tumours, bacterial, fungal or viral infections, depression and

CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given  
CC in the exemplification of the present invention.

XX Sequence 437 BP; 119 A; 96 C; 101 G; 121 T; 0 other;

Query Match 17.4%; Score 408.6; DB 21; Length 437;  
Best Local Similarity 99.0%; Pred. No. 2.7e-107;  
Matches 411; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1096 GCTTATGAAAGTACCTCTGCTGATGCTGAACTTGTGCAAGTCTATTTTC 1155

Db 21 GCTTATGAAAGTACCTCTGCTGATGCTGAACTTGTGCAAGTCTATTTTC 80

Qy 1156 AGTGTGGTTTAAAGCTATTAATGCTGCGAGTCAAGGAATGTGGCAAGTACCTT 1215

Db 81 AGTGTGGTTTAAAGCTATTAATGCTGCGAGTCAAGGAATGTGGCAAGTACCTT 140

Qy 1216 ACTGTTGAGATCTAGTAATGCTGAAAGTACCTTTTTCAGCTTTTCAATTAACCTTGAAC 1275

Db 141 ACTGTTGAGATCTAGTAATGCTGAAAGTACCTTTTTCAGCTTTTCAATTAACCTTGAAC 200

Qy 1276 TTTCTGGGAACCTGATATGAGAGACTAGACAGCACTCATGATATGAACCTTTGTT 1335

Db 201 TTTCTGGGAACCTGATATGAGAGACTAGACAGCACTCATGATATGAACCTTTGTT 260

Qy 1336 ACTCTACTCAAGGTAGACACCCAAATTAAGACAAAGTATGAGCATCTCCCTTCAGATC 1395

Db 261 ACTCTACTCAAGGTAGACACCCAAATTAAGACAAAGTATGAGCATCTCCCTTCAGATC 320

Qy 1396 ACACCAAGACAGCTACCGCTGCTTTGATTAATGCTGATTTTGAATATCATTTAGAGGCCAG 1455

Db 321 ACACCAAGACAGCTACCGCTGCTTTGATTAATGCTGATTTTGAATATCATTTAGAGGCCAG 380

Qy 1456 AAAGTCTTATGAGTAATCTTTGAAGTCCCTGACAGGAAAGAAATGCGCCATTG 1510

Db 381 AAAGTCTTATGAGTAATCTTTGAAGTCCCTGACAGGAAAGAAAGTCCGCTCG 435

## RESULT 12

AAx78080  
ID AAx78080 standard; DNA; 5369 BP.

AAx78080;

19-AUG-1999 (first entry)

DE A. goseypil ADE4 DNA.

XX Purine biosynthesis; phosphoribosylpyrophosphate synthetase; KPR1; KPR2;

KW glutamine-phosphoribosylpyrophosphate amidotransferase; ADE4; GUA1;

KW IMP-dehydrogenase; GUA1; GMP-synthetase; GUA2; PRPP synthetase; mouth;

KW riboflavin; vitamin B2; treatment; inflammation; throat; mucosal layer;

KW skin disorder; ss.

XX Ashbya goseypil.

OS

XX Key

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

XX 23-DEC-1997; 97DE-1057755.

PR (BADI) BASF AG.

XX Hoeftlen HW, Jimenez A, Pompejus M, Revuelta Doval JL;

PI Santos Garcia MA, Seuburger H;

XX WPI: 1999-373465/32.

DR P-PSDB; AA08961, AA08962, AA08963.

XX Ashbya goseypil purine biosynthesis enzymes and related genes for

FT production of riboflavin

PS Claim 12; Page 15-22; 48pp; German.

CC This invention describes novel enzymes involved in purine biosynthesis

CC in Ashbya goseypil. The enzymes described in the invention include

CC phosphoribosylpyrophosphate (PRPP) synthetases KPR1 and KPR2

CC glutamine-phosphoribosylpyrophosphate amidotransferase (ADE4)

CC IMP-dehydrogenase (GUA1) and a GMP-synthetase (GUA2). The nucleic acid

CC sequences encoding the enzymes involved in purine biosynthesis are used

CC to genetically modify microorganisms, for the production of riboflavin.

CC Riboflavin, also known as vitamin B2, is essential in humans and animals

CC as well as used to treat inflammation of the mouth and throat mucosal layer

CC ADE4 gene which encodes a glutamine PRPP amidotransferase and 2 other

CC open reading frames.

XX Sequence 5369 BP; 1351 A; 1333 C; 1299 G; 1386 T; 0 other;

Query Match 15.3%; Score 358.6; DB 20; Length 5369;

Best Local Similarity 54.1%; Pred. No. 3.1e-97;

Matches 777; Conservative 0; Mismatches 69; Indels 9; Gaps 2;

Qy 703 AACCTGATCTGGGTTTACCTGACAGACAGACGAGCTTATCTAAGCTATTGAC 762

Db 1 AAGCTTGACCTTGGCTGACCTGATGCGACAGAGTGAATCAACCGCAATGAT 60

Qy 763 AGAGCAACAGGGATACGTTTCTGCTGAGTGTGATTAATCTTCCCATC 822

Db 61 CGTGTGTAAAGATCTCTTAATGCTGACAGTCAATGTTTTCATATATCCGATT 120

Qy 823 ATGTTGAAGTATGCTGCTGAGTGTGTTTGTATTAACAATCGGTGCCAGTTGCT 882

Db 121 ACATTGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180

Qy 883 TTGTAACCTTGGAAACCTGTTGATACATACACAGCATTCACAGTTGACATCGGTG 942

Db 181 GCTATTAACATTCGACATATCTCTTACATCACTTTACTTTACAGAACGACGCGTGG 240

Qy 943 AGAAGTATTAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1002

Db 241 CGACACGCTTTAGCGGTGATGACGACGACGACGACGACGACGACGACGACGACGAC 300

Qy 1003 GACTCAGTCTGAATATGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1062

Db 301 GATTCCTTAATTAATTTTGAAGCTGTAAGATTTTCAATTAACGAACTACCTGGGAC 360

Qy 1063 AGATATGATGATTTTGAAGACGTATGACGCTCTTATGAAAGTACTTACTCTG 1122

Db 361 AAGATATGACATCTCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 420

Qy 1123 GCTATGCTGAATTTTGTCAAGTGTATTTTGTGCTGCTTTTAAAGCTATATGATG 1182

Db 421 GCGTTTGAACACCGGCGACGACGACGACGACGACGACGACGACGACGACGACGAC 480

Qy 1183 CTCGCAAGTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1242

Db 481 ATGCTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540

Qy 1243 GAGTCTTTTCAAGCTTTCAATCCCTGAACTTTCTGGAACTGTATATGAGAGACT 1302

Db 541 CAACCTGATTCACAGCTCCGTGCACTAACTTCCTGTTAGCGTCTACCGTATCTC 600  
Qy 1303 AGACAGCACTCATAGATATGACACCTTTGTTACTCTACTCAAGTAGACACCCAAATT 1362  
Db 601 AAGAGCTCTGATAGATATGAAATCTTTATTTAACTGCAAAAAATCAGSTCAAAATT 660  
Qy 1363 AAAGACAAAGTATGGCATCTCCCTTCAGATCAACACAGACGACTACCGTGCCTTT 1422  
Db 661 AAGAACTCCCAAAATGCCAGAACCTACCAATACAAACCGTTGATAT---TTGCTTT 717  
Qy 1423 GATATATGATTTTGAATATATGAGGGCCAGAAAGCTCTTAATGGAATTCCTTGAA 1482  
Db 718 GAATATGATATGATTTGCTATGACCCGAGCGGTATATGAACAATGTTTCTTACC 777  
Qy 1483 GTCCCTCAGAGAAAGAAAGTGGCCATTTGAGAGATAGTGGGTGAGGAAAGACAAATA 1542  
Db 778 ATCCAGCTGGAATGAAGACGTCCAGATAGCCCATCGGGCTCGGGGAAGTCCACCAATT 837  
Qy 1543 GTGAGCTATTTATTCGCTTCTATGAGCTTCAAAAGGTAGCAATTTATCTTGCTGTCAA 1602  
Db 838 TTGAAGCTCGATTTAGATTTCTATGAGCCGAGCAAGTCTATCTATGCGGACACA 897  
Qy 1603 AATATACAAATGATGAGCTTGAAAGCTTGGAGGCGAGTGGAGTGTACTCAGAT 1662  
Db 898 GATATCCGCGATTTAGACTTGCTTTCTTACGAGAGCTATCGGTGCTGCCCAAGAT 957  
Qy 1663 GCTGCTCTCTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1722  
Db 958 ACTGCTCTCTTCAATGACAACTCGGAGAAATGTTAAATTCGGCAATATATAGTCTCT 1017  
Qy 1723 CTTGAGGAAGTATGACAGTGGCAAAATTAAGTGAATTCATGATGCAATTCCTTGAATG 1782  
Db 1018 GACGATGATTTCTCAGGGCCATGAAAAAGCTCACTCAGAAAGTACTCAGAACCTA 1077  
Qy 1783 CCAATGATATGACACCCCAAGTGGGGAACGAGACCTCAAGCTTTAGAGGAGAAAG 1842  
Db 1078 CCAAAAGGCGCTTCACCGTTGTGGGAGCGCGTTGATGATCAGCGGAGTGAGAAA 1137  
Qy 1843 CAAAGATGAGCAATTCGAAAGACCATTTTGAAGAGCCCGGAGTCACTACTATATGATA 1902  
Db 1138 CAAAGGCTTGATTTGCTCGTGTGCTTTTGAAGACGCTCCGCTGATGTTTTCAGAG 1197  
Qy 1903 GCTACTCATGTTAGATTGCTGATTACTGAAAGACTATTTGTTGGTGCAT-----GAAG 1956  
Db 1198 GCTACAGTGTCTGATACACACAGAGGACGACTTTCGACACATTCACAGAGAC 1257  
Qy 1957 GATGTGATCAACACAGAACTTCTATTTTCATTCGACACAGATTTCAACAGTGTGAT 2016  
Db 1258 TTTTCTTCAATTCAAAGACGAGCGTTTACGTTCCCATGACCTGCCACATGCTGAT 1317  
Qy 2017 GCAGATGAATCATTTGCTTGATCAGGGTTAAGGTAGCCGAGGTGTAACCAACATGAT 2076  
Db 1318 GCAGATGAATCATTTGCTTGATCAGGGTTTCTTTCGCGAGAGGGGACACACAGCTCG 1377  
Qy 2077 TTGCTTCAACCTCATATGATTTATTCAGAAATGTGATACACAGAGAGCC 2131  
Db 1378 CTGTTAGCGTCAAGAGATCCCTATACCGGGGCTGTGGGATATTCAGAAAAACC 1432

## RESULT 13

AAC37701

ID AAC37701 standard; DNA; 1185 BP.

XX AAC37701;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 18355.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX

XX

PN EP1033405-A2.

XX

XX

PD 06-SEP-2000.

XX

XX

PE 25-FEB-2000; 2000EP-0301439.

XX

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135533.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137528.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144337.  
PR 19-JUL-1999; 99US-0144337.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144864.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.

PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 13.5%; Score 316.6; DB 21; Length 1185;  
Best Local Similarity 58.4%; Pred. No. 1,7e-80;  
Matches 594; Conservative 0; Mismatches 414; Indels 9; Gaps 2;

QY 1132 AACTTGGTCMAAGTGCATTTTCAGTGTGGTTTAAACGCTAATATGTCGCCAGT 1191  
DB 1 AACTTGGACMAAGTATCATTTAGTACAGCTTTATCAACGCAATGTTGTGCTCT 60  
QY 1192 CAGGAATGCGCAGGTACCTTACTGTGGAGATCATGATTAAGTGAATGAGTGCCTT 1251  
DB 61 CAGGTATCATGAATGCGCAGATACAGTGGTATTTGGTATGATGATGACTTCTC 120  
QY 1252 TTTGAGCTTTCATTAACCCCTGAACCTTCTGGAACTGATATAGAGACTAGACA 1311  
DB 121 TTTGAGCTGTCTTCCCTCTAACTTTTGGTAGTGTATTCGGTGAACATTCAGAGC 180  
QY 1312 CTCATGATATGAACACCTTGTCTTACTCTACTCAAGTAGACACCCAAATTAAGACAA 1371  
DB 181 CTTGTGACATGAATCAATCATTTCCAGTTGCTAGAGGAATCTGACATCACAAAATACA 240  
QY 1372 GTGATGCAATCTCCCTTCAGATCAACACACAGACGTAACCGGCTTGAATATG 1491  
DB 241 AGTAGCGAAAGCCTCTTGAATGAAGGT-----GAAACATTAAGTTGAAGAAATGA 294  
QY 1492 CATTTGAATACATTTGAGGGCCAGAAAGTCTTAGTGAATATCTTTGAAGTCCCTGA 1491  
DB 295 CACTTCAGTTACCTTCCAGAGAGAAATATATTAATGGAATTTCTTTGTCACCGCA 354  
QY 1492 GGAAGAAAGTGCCCATTTGATGAGAGTAGTGGTCAGAGAAACACAAATGAGGCTA 1551





```

Db 1783 ACAGAAAGAACTTCGTGACATGAGAGAACTGTCATCTGCGAAAGAGAGAGAGAT 1842
Oy 1362 TAAAGCAAAAGATGATGATCTCCCTTCAGATCAACACAGACAGACGACGCGCTT 1421
Db 1843 TGTGATGCGCCAGGCTGTTGGCCCTTCTTAACCTGCC-----GGCGGGGCGATTGAGTT 1896
Oy 1422 TGATATATGATGATTTGATATGATGAGGGCCAGAAAGCTTATGATGAAATATCTTTGA 1481
Db 1897 CTCAAAAGCTGACCTTTGGCTATTCACCGAGAAAGATAGTCTGCCCAATGTAGCTTAC 1956
Oy 1482 AGTCCCTGAGGAAAGAAAGTGGCCATTTGAGAGGTAGTGGTACGAGAAAGACAAAT 1541
Db 1957 AGTCCGCGCGGTAGACGCGTAGCATGTGTGACCTCTGAGCGCGGTAAAGACACAT 2016
Oy 1542 AGTGGAGCTATATTTCCGTTCTATGAGCGCTCAAAAGGTATGACATTTATCTTCTGCTCA 1601
Db 2017 CATGGCGGCTGCTTTTCCGCTTTTACGATGTGACACTGTGTATCTGATGACGAGCA 2076
Oy 1602 AAATATACAAAGATGATGAGCGCTGAGAAAGCTTGGAGGAGTGGAGTGGTACTCAGGA 1661
Db 2077 AAACATCAAGCTTGTGACGACAGACAGTCTGCGAAAGGCCATTGAGTGTGCTCCCGCAGGA 2136
Oy 1662 TGCTGTCTCTTCCATATATATATATATTAACCTCTTATATGAAACATGAGTCTTC 1721
Db 2137 TACGGTCTCTTAAACAACACCATCTTCAACAATGAAATGACCAAGTTGGGCGCTTC 2196
Oy 1722 ACTTGAGAGAGTGTATGACGAGTGGCAAAATTAAGTGAATGATGATGATGATGATGAT 1781
Db 2197 GGACGAGAGGTGTCTATGAGGCTGCTCGCGCCCGCAGCATTCAGAAAGGATTTCTGGCTT 2256
Oy 1782 GGCACATGATATGACACCCAGTAGAGGGAAGAGAGTCAAGCTTTTCAGAGAGAGAAA 1841
Db 2257 TCCGGAAGAAATATGAGACAAAGTGGGCGAAGAGGTCTCGGCTCAGTGGAGGTGAAA 2316
Oy 1842 GCAAGAGTAGCAATTTGCAAGAGCCATTTTGAAGAACCCCGCAGTCAATCTATGATGA 1901
Db 2317 GCAAGCGGTAGCATGCGCAGAAAGCTTCTTAAGGCGCCATTAATGTGCTCCTGAGATGA 2376
Oy 1902 AGCTACTTCATGCTAGATTCATTAAGAGACTATCTTGTGGCTCAGTAAGAGATGT 1961
Db 2377 AGCCACCTTAGCTTTGAGACCCACAGACAGACGATATATTCAGGCTGCTTGGCAGAGGT 2436
Oy 1962 GGTCAACACAGAACTTCTATTTTATGTCACACAGATTTCAAGAGTGGTGTGATGACGA 2021
Db 2437 GTGCCCAATGCTACACCATATATGTTGGCCATGTCTTTCCACCATCATCATGCGGA 2496
Oy 2022 TGAATTCATTTGCTTGGATCAGGGTAAAGGTAGCCGAAAGCTGTGTACCCACCATGTGTTGCT 2081
Db 2497 CGAGATTTCTGTCGTGACAGCAGGAAAGCATAGCGAGCGGTGGCGCAGAGAGACTTGT 2556
Oy 2082 TGCTAACCTCATATGATATATTCAGAAATGTGGCA 2117
Db 2557 GCTGGCGCAAGATGCATCTACGCGACATGTGGCA 2592

```

## RESULT 15

AAD27254  
ID AAD27254 standard; cDNA, 2984 BP.

AC AAD27254;

DT 09-APR-2002 (first entry)

XX Human transporter and ion channel-1 (TRICH-1) cDNA.

XX Human; transporter and ion channel-1; TRICH-1; neuroprotective; asthma;  
XX noctropic; cytosolic; cardiovascular; immunosuppressive; cardiomyopathy;  
XX antiinflammatory; protein therapy; akinesia; cystic fibrosis; leukaemia;  
XX Bell's palsy; amyotrophic lateral sclerosis; Alzheimer's disease; cancer;  
XX amnesia; dementia; myocardiitis; Duchenne's muscular dystrophy; AIDS;  
XX Acquired Immune Deficiency Syndrome; Addison's disease; allergy; angina;  
XX cell proliferative disorder; psoriasis; cardiac disease; hypertension;

```

KW bradyarrhythmia; gene expression; drug screening; ss.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH 269..2797
FT /tag= a
FT CDS /product= "Human TRICH-1 protein"
FN WO200192304-A2.
PD
PF 06-DEC-2001.
PF 25-MAY-2001; 2001MO-US17065.
XX
PR 26-MAY-2000; 2000US-208424P.
PR 01-JUN-2000; 2000US-209001P.
PR 08-JUN-2000; 2000US-210588P.
PR 16-JUN-2000; 2000US-212335P.
PR 22-JUN-2000; 2000US-213747P.
PR 29-JUN-2000; 2000US-215391P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Thornton M, Walla NK, Yue H, Nguyen DB, Lai P, Gandhi AR;
PI Triboley CM, Yao MG, Ramkumar J, Au-Young J, Lu Y, Tang YT;
PI Azimzai Y, Bruns CM, Griffin JA, Yang J, Sanjambala MS;
PI Raumann BE, Lee EA, Hafalia A, Baughn MR, Green BD, Khan FA;
PI Kearney L, Elliot VS, Seilhammer JU, Policky JU, Borowsky ML;
PI Burford N, Ding L, Lu DM, Hallman JL;
XX
DR WPI: 2002-122055/16.
DR P-PSDB; AAE16764.
XX
PT New human transporters and ion channels (TRICH) polypeptides useful for
PT diagnosing, treating or preventing disorders associated with aberrant
PT expression of TRICH -
PT
PS Claim 11; Page 187-188; 210pp; English.
XX
CC The invention relates to human transporters and ion channels (TRICH)
CC polypeptides and their cDNA molecules. The nucleic acid and polypeptide
CC sequences are useful in the diagnosis, treatment, and prevention of
CC disorders associated with transport (akinesia, cystic fibrosis, Bell's
CC palsy, amyotrophic lateral sclerosis); neurological (Alzheimer's disease,
CC amnesia, dementia); muscle (cardiomyopathy, myocardiitis, Duchenne's
CC muscular dystrophy); immunological (AIDS, Addison's disease, allergies,
CC asthma); cell proliferative disorders (cancers, leukaemia, psoriasis);
CC cardiac disease (angina, hypertension, or bradyarrhythmia) and in the
CC assessment of the effects of exogenous compounds on the expression of
CC nucleic acid and amino acid sequences of transporters and ion channels.
CC The polynucleotides may be used to detect and quantify gene expression
CC in biopsied tissues in which TRICH expression may be correlated with a
CC disease, to generate hybridization probes for mapping naturally occurring
CC genomic sequence, and in drug screening. The present sequence is human
CC TRICH-1 cDNA.
XX
SQ Sequence 2984 BP; 571 A; 831 C; 916 G; 666 T; 0 other;

```

Query Match 12.1%; Score 284.2; DB 24; Length 2984;

Best Local Similarity 50.6%; Pred. No. 6,4e-71;

Matches 744; Conservative 0; Mismatches 718; Indels 9; Gaps 2;

```

Oy 666 TCTTTTCATCTTCACAACTGATCTGGTTTTCACCTGAGACAGACAGCGGAGCTT 745
Db 1314 TCTTCTCCACCTCAGACAGCTCTCAGCTGCGTGGACCTGGCGCGCAGAGGAGG 1373
Oy 746 TATCAAGGATATGACAGAGAGAGGATATCAGTTTGTCTGAGTCTTTGAT 805
Db 1374 TGCTGGGATGGCGATGGGCGACATCCATGTCACAGGCTCTCAGCTAGTGGT 1433
Oy 806 TTAATCTTCTTCCATCATGTTTGAAGTATGCTTGTGAGTGTGTTGTATTAACAAT 865

```

Db 1434 TCAATGATCCCAAGCTGGCCGACATCATCTTGGCATCTACTTCAAGATGTTCT 1493  
 QY 866 GCGGTCCCAAGTTGCTTTGGTAACTTGGAACTTGGATACACAGCATTTCAAG 925  
 Db 1494 TCAACGCTGGTGGCTTCAATGTTGTTCTGTGATGAGCTTTTACTTCACTCCGACCA 1553  
 QY 926 TTGCAATCAACGGTGGAACTAGATTAGAAATGAAATGAAACAAAGCAGATTAATG 985  
 Db 1554 TTGGGTCACTGAGTGGAAACCAAGTTTCGTGCTATTAACAACAGAGAAACGCTA 1613  
 QY 986 CAGGTAATGCTGCTATAGACTCACTGCTGAATTAATGAACGTGAAGTATTAATATG 1045  
 Db 1614 CCGGGACAGACAGAGTGAATCTTCTGTAACCTTGAACGATGAAGTATTAACAACCG 1673  
 QY 1046 AAAGATATGAAGCAAGAGATATGATGATTTTGAACGCTATGAGACTGCTTCAATGA 1105  
 Db 1674 AGAGTTACGAAGTGAACCGCTATCGAAGGCCATCATCAATATCAAGGCTTTGAAGTGA 1733  
 QY 1106 AAAGTACTCTACTCTGCTATGCTGAATTTGGTCAAAAGTCTAATTTTCAAGTGTG 1165  
 Db 1734 AGTGAAGCGCTTCACTGTTTACTTAATCAGACCCAGAACCTGTGATTTGGGCTCG 1793  
 QY 1166 TAAAGCTATTAATGCTGCTGCGCAGTCAAGGAAATTTGGCAGTACCTTACTGTTGAG 1225  
 Db 1794 TCCCTGCGCGCTCCCTGCTTTGGCCATATCTTTGCACTGAGCAGAGCTTACAGGTTG 1853  
 QY 1226 ATCTAGTAATGATGATGAAGTGAATGCTTTTTCAGCTTTTCAATTAACCTGAACTT 1285  
 Db 1854 ACTATGCTCTTTGGGACCTACATTAATCAAGCTGTATCATGCTCCCTCACTGCTTTG 1913  
 QY 1286 CTGATATTAAGAGACTAGACAGACACTCATATATGAACACTTGTCTTACTCA 1345  
 Db 1914 CTTACTAGAGATATCAGACCAACTTCATTTGACATGAGAACATGTTGACTTGTCTGA 1973  
 QY 1346 AGGTAGACACCCAAATTAAGAAAGATGATGATCTCCCTTCAATCAACACACAGA 1405  
 Db 1974 AAGGAGACAGAAATGAGAGCTTCTCTGAGCAGGCGCTTCTGCTT-----CAGA 2027  
 QY 1406 CAGCTACCTGCTGCTTGAATATGATGATTTTGAATACATTTGAGGAGCAGAAAGTCTTA 1465  
 Db 2028 AGGCGCGATTAATGATTTGAGAGAGTGAACCTTCACTATGCGATGGGAGAGACTGTC 2087  
 QY 1466 GTGAAATATCTTTGAAGTCCCTGCAAGAAAGATGGCCATTTGAGAGTATGAGGT 1525  
 Db 2088 AGGACGTCTTTCACTGTATGCTGGAACAGACACTTGCCTGTGGGCCCATCTGGGG 2147  
 QY 1526 CAGGAAAGACAAATAGTGGCTATTAATTCGCTTCTATGAGCTCAAAAAGGTAGCA 1585  
 Db 2148 CAGGAAAGACAAATTTGGCTGTGCTTGTCTTCAAGACATGAGCTTGGCTGCA 2207  
 QY 1586 TTTATCTTGTGCTCAAAATATCAAGATGAGCTGAAAGGCTTCGAGGAGCAGTGG 1645  
 Db 2208 TCCGAATAGATGGCAGAGACTTTCAAGGTGACCCAGGCTCTCTCCGCTTCAATTTG 2267  
 QY 1646 GAGTGTACTCTGAGATGCTGTCTCTTCAATAATTAATTAATTAATTAATTAATG 1705  
 Db 2268 GAGTTGTGCCCCAAGACACTGTCTCTTAATGACACATGCGCAGACATATCGTTAAG 2327  
 QY 1706 GAAACATCAGTGTCTTCACTGAGAGATGATGAGTGGCAAAATTAAGTGGACTTATG 1765  
 Db 2328 GCCGTGTCAAGCTGGAAATGATGAGGTGCTGCTCAGGCTGCAAGGCATTCATG 2387  
 QY 1766 ATGCAATTTCTTGAATGCTCAATGATATGACCCAAAGTAGGGGAGAGAGCTCAAGC 1825  
 Db 2388 ATGCAATTAATGCTTTCTCTGAGAGGTATCAGAGACAGAGTGGGAGAGGAGCTGAGC 2447  
 QY 1826 TTTCAAGAGAGAAAGCAAAAGATAGCAATTGCAAGAGCCATTTTGAAGAACCCCCAG 1885  
 Db 2448 TGAAGCGGCGGGAGAAAGCAGCGCTGCAATTTGCGCCAGCAGATCTCAAGGCTCCGGCA 2507  
 QY 1886 TCAATCTATGATGAGTACTTCAATGTTAATTTGATTTACTGAGAGAGATATTTCTTG 1945  
 Db 2508 TCATTTGCTGATGAGCAACGTCAGGCTGATACATCTAATGAGAGGCGCATTCAGG 2567

QY 1946 GTGCCATGAAGATGTGTCAAAACAGAACTTCTAATTTTCAATGACACAGATTTGCAA 2005  
 Db 2568 CTTCCTGGCCAAAGTGTGTGCCAACCGCACACACATCGTAGTGGACACAGGCTCTCAA 2627  
 QY 2006 CAGTGTGATGACAGATGAATATTTGTTGATCAGGGTAAAGTATCCGAAAGTGTGA 2065  
 Db 2628 CTGTGTCAATGCTGACAGATCTCTGTATCAAGATGAGCTGATCTGTGAAGAGGAGAC 2687  
 QY 2066 CCACCATGTTGCTTGTCTAACCTCATAGTATCTATTGAGAAATGTGGCATTACAGA 2125  
 Db 2688 GACACAGGCTCTGTGTCT---CCGAGTGGGTGTATGCTGACATGTGGACAGCTGAGC 2744  
 QY 2126 GCAGCCGTGTGCAAAACCATGATTAACCCCA 2156  
 Db 2745 AGGAGACAGAAAGAACTCTGAAAGACCTTA 2775

Search completed: February 15, 2003, 01:40:17  
 Job time : 512 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2003, 01:30:25 ; Search time 87 Seconds  
(without alignments)  
8266.172 Million cell updates/sec

Title: AF133659  
Perfect score: 2345  
Sequence: 1 ATGCGCGCTGCTGCGATGCA.....TATTAAAAAATCATACATT 2345

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/prodata/2/ina/5A COMB. seq: \*  
2: /cgn2\_6/prodata/2/ina/5B COMB. seq: \*  
3: /cgn2\_6/prodata/2/ina/6A COMB. seq: \*  
4: /cgn2\_6/prodata/2/ina/6B COMB. seq: \*  
5: /cgn2\_6/prodata/2/ina/PCITUS COMB. seq: \*  
6: /cgn2\_6/prodata/2/ina/bcktitles1. seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 2248  | 95.9        | 2407   | 2  | US-08-895-522-2     |
| 2          | 2248  | 95.9        | 2407   | 3  | US-09-195-591-2     |
| 3          | 358.6 | 15.3        | 5369   | 4  | US-09-212-247C-3    |
| 4          | 188.8 | 8.1         | 1749   | 4  | US-09-134-001C-1893 |
| 5          | 180.2 | 7.7         | 7760   | 4  | US-08-961-527-63    |
| 6          | 170   | 7.2         | 5120   | 3  | US-08-772-370A-6    |
| 7          | 170   | 7.2         | 8370   | 2  | US-08-488-706-1     |
| 8          | 165.8 | 7.1         | 13188  | 4  | US-08-961-527-70    |
| 9          | 165.4 | 7.1         | 3576   | 4  | US-08-976-259-79    |
| 10         | 154.8 | 6.6         | 6452   | 4  | US-08-961-527-188   |
| 11         | 147   | 6.3         | 7147   | 4  | US-08-961-527-23    |
| 12         | 143.8 | 6.1         | 2726   | 1  | US-08-461-823-1     |
| 13         | 143.8 | 6.1         | 4264   | 2  | US-08-784-649A-1    |
| 14         | 143.8 | 6.1         | 4264   | 2  | US-08-784-649A-5    |
| 15         | 143.8 | 6.1         | 4664   | 1  | US-08-181-471-2     |
| 16         | 143.8 | 6.1         | 4669   | 6  | 5206352-3           |
| 17         | 143.8 | 6.1         | 6505   | 2  | US-08-793-610-5     |
| 18         | 143.8 | 6.1         | 9318   | 2  | US-08-793-610-6     |
| 19         | 142.2 | 6.1         | 4669   | 2  | US-08-583-276-18    |
| 20         | 140.6 | 6.0         | 4669   | 2  | US-08-752-447-1     |
| 21         | 139.4 | 5.9         | 1201   | 3  | US-09-316-167-1     |
| 22         | 139.4 | 5.9         | 1201   | 3  | US-08-961-083-189   |
| 23         | 139.4 | 5.9         | 7186   | 4  | US-08-961-527-39    |
| 24         | 135.2 | 5.8         | 1743   | 4  | US-09-134-001C-2774 |
| 25         | 134.8 | 5.7         | 11864  | 4  | US-08-961-527-61    |
| 26         | 131   | 5.6         | 7721   | 3  | US-08-772-370A-14   |
| 27         | 129.4 | 5.5         | 4233   | 3  | US-09-120-513-1     |

|    |       |     |         |   |                    |                    |
|----|-------|-----|---------|---|--------------------|--------------------|
| 28 | 129.4 | 5.5 | 4233    | 4 | US-09-450-105-1    | Sequence 1, Appl1  |
| 29 | 126   | 5.4 | 775     | 4 | US-09-221-017B-860 | Sequence 860, App  |
| 30 | 125.8 | 5.4 | 4002    | 2 | US-08-996-545-1    | Sequence 1, Appl1  |
| 31 | 125.8 | 5.4 | 4002    | 2 | US-08-996-545-3    | Sequence 3, Appl1  |
| 32 | 125.8 | 5.4 | 4002    | 4 | US-09-328-320-1    | Sequence 1, Appl1  |
| 33 | 125.8 | 5.4 | 4002    | 4 | US-09-328-320-3    | Sequence 3, Appl1  |
| 34 | 125.2 | 5.3 | 4047    | 2 | US-08-612-734B-1   | Sequence 1, Appl1  |
| 35 | 125.2 | 5.3 | 4800    | 2 | US-08-612-734B-3   | Sequence 3, Appl1  |
| 36 | 116.6 | 5.0 | 28473   | 4 | US-08-612-527-83   | Sequence 83, Appl1 |
| 37 | 116.2 | 5.0 | 4224    | 1 | US-08-612-521-1    | Sequence 1, Appl1  |
| 38 | 112.6 | 4.8 | 1749    | 4 | US-09-221-017B-703 | Sequence 703, App  |
| 39 | 109   | 4.6 | 14672   | 4 | US-08-961-527-111  | Sequence 111, App  |
| 40 | 108.8 | 4.6 | 5595    | 1 | US-07-841-655-1    | Sequence 1, Appl1  |
| 41 | 108.8 | 4.6 | 5595    | 1 | US-07-635-965C-1   | Sequence 1, Appl1  |
| 42 | 108.6 | 4.6 | 3909    | 1 | US-08-232-537-1    | Sequence 1, Appl1  |
| 43 | 107.8 | 4.6 | 1971    | 4 | US-08-858-207A-72  | Sequence 72, Appl1 |
| 44 | 107.8 | 4.6 | 4403765 | 4 | US-09-103-840A-2   | Sequence 2, Appl1  |
| 45 | 107.6 | 4.6 | 3924    | 2 | US-08-996-644-3    | Sequence 3, Appl1  |

## ALIGNMENTS

RESULT 1  
US-08-895-522-2  
; Sequence 2, Application US/08895522  
; Patent No. 5858719  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Shah, Purni  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE  
; TITLE OF INVENTION: TRANSPORT PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Inocyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/895,522  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0336 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2407 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: OVARNOT02  
; CLONE: 545981  
; US-08-895-522-2  
Query Match 95.9%; Score 2248; DB 2; Length 2407;





QY 2168 AGAAGAAATATATCCAAAGAGAGAGAAAGAACTACAAGAAATTTGCAATA 2227  
| | | | |  
Db 2161 AGAAGAAATATATCCAAAGAGAGAGAAAGAACTACAAGAAATTTGCAATA 2220  
| | | | |  
QY 2228 GTGGAAGGCTGTGGAAAGCTTTCTGCTAAGTCATAGACATTTCTTTTGT 2286  
| | | | |  
Db 2221 GTGGAAGGCTGTGGAAAGCTTTCTGCT-AGTCACATAGACATTTCTTTTGT 2278  
| | | | |

## RESULT 2

US-09-195-391-2  
; Sequence 2, Application US/09195391  
; Patent No. 6080842  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Shah, Puri  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE  
; TITLE OF INVENTION: TRANSPORT PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/195,391  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/895,522  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0336 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2407 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: OVARNOF02  
; CLONE: 545981  
; US-09-195-391-2

Query Match 95.9%; Score 2248; DB 3; Length 2407;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 2265; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 8 TGTCTGGGATGATCTTTGGGCGCTGGCGCGCGCGCGCTTTGGAAAAGCGCGGC 67  
| | | | |  
Db 1 TGTCTGGGATGATCTTTGGGCGCTGGCGCGCGCGCGCTTTGGAAAAGCGCGGC 60  
| | | | |  
QY 68 ACTCCGGATTTCTGATCCGGCGCTTAACTCTCTTTAGCGGCTCAGTCCGAGTGAAGGC 127  
| | | | |  
Db 61 ACTCCGGATTTCTGATCCGGCGCTTAACTCTCTTTAGCGGCTCAGTCCGAGTGAAGGC 120  
| | | | |  
QY 128 CACATCAACTCGCGGCTTGGGAAACCGCTGAGGCTACCAAGTTCCAGATCATTAATAA 187  
| | | | |

Db 121 CACATCAACTCGCGGCTTGGGAAACCGCTGAGGCTACCAAGTTCCAGATCATTAATAA 180  
| | | | |  
QY 188 GTATCACATGGCAGAGATTGGGAAAAGGCAATTCAGGACAGTTCTTATGATCTGCAAGG 247  
| | | | |  
Db 181 GTATCACATGGCAGAGATTGGGAAAAGGCAATTCAGGACAGTTCTTATGATCTGCAAGG 240  
| | | | |  
QY 248 CTCTCCAGTATGGCCACTGATAGAAAAGAGACATGTTGTCATATGCAAGAGAG 307  
| | | | |  
Db 241 CTCTCCAGTATGGCCACTGATAGAAAAGAGACATGTTGTCATATGCAAGAGAG 300  
| | | | |  
QY 308 GACTCCACACAGACCCAAAAGAGGGTTAAAAGATGTTGATCTCGGAAAATCATTAAG 367  
| | | | |  
Db 301 GACTCCACACAGACCCAAAAGAGGGTTAAAAGATGTTGATCTCGGAAAATCATTAAG 360  
| | | | |  
QY 368 CAATGCTTTCTTATGTGTGGCCCAAGAGCGCCAGATCTACGAGCTAGAGTCCCATTT 427  
| | | | |  
Db 361 CAATGCTTTCTTATGTGTGGCCCAAGAGCGCCAGATCTACGAGCTAGAGTCCCATTT 420  
| | | | |  
QY 428 CGCTGGGATTTTGGGTGTGCAAAAGCCATGATATTTGTTCCCTTCATGTTTAAAT 487  
| | | | |  
Db 421 CGCTGGGATTTTGGGTGTGCAAAAGCCATGATATTTGTTCCCTTCATGTTTAAAT 480  
| | | | |  
QY 488 ATGCTGTAGACAGCTTACCAAGATGTGCGGAAAACATGCTGAACCTGATGACACCA 547  
| | | | |  
Db 481 ATGCTGTAGACAGCTTACCAAGATGTGCGGAAAACATGCTGAACCTGATGATCACCA 540  
| | | | |  
QY 548 ATACAGTTGCAACCATGAGCAACAGCAGTTCGATGTTGCTATGCTATCAAGAGCTGAG 607  
| | | | |  
Db 541 ATACAGTTGCAACCATGAGCAACAGCAGTTCGATGTTGCTATGCTATCAAGAGCTGAG 600  
| | | | |  
QY 608 CTGCTTTTAAAAAGAGTTGCAAAATGCAATTTGGCAAGTATGCCAGAAATCAATCC 667  
| | | | |  
Db 601 CTGCTTTTAAAAAGAGTTGCAAAATGCAATTTGGCAAGTATGCCAGAAATCAATCC 660  
| | | | |  
QY 668 GAAAGATAGCCAAAATGCTTTCTCATCTTCAACCTGATCTGGGTTTCACTGA 727  
| | | | |  
Db 661 GAAAGATAGCCAAAATGCTTTCTCATCTTCAACCTGATCTGGGTTTCACTGA 720  
| | | | |  
QY 728 GCAGACAGAGGGAGCTTATCTAAGGCTATGACAGAGAACAGGGGATACGTTTG 787  
| | | | |  
Db 721 GCAGACAGAGGGAGCTTATCTAAGGCTATGACAGAGAACAGGGGATACGTTTG 780  
| | | | |  
QY 788 TCTGAGTCTTTGATTAATCTTCTCCATCATGTTTGAAGTATGCTGTCAGTG 847  
| | | | |  
Db 781 TCTGAGTCTTTGATTAATCTTCTCCATCATGTTTGAAGTATGCTGTCAGTG 840  
| | | | |  
QY 848 GTGTTTGTATTAACAATGCGGTGCCAGTTTGCTTGTATACCTTGAACCTTGTA 907  
| | | | |  
Db 841 GTGTTTGTATTAACAATGCGGTGCCAGTTTGCTTGTGTATACCTTGAACCTTGTA 900  
| | | | |  
QY 908 CATACACAGATTCACAGTTGCGAGTCAACAGGTGAGAACTAATTTAGAAATGA 967  
| | | | |  
Db 901 CATACACAGATTCACAGTTGCGAGTCAACAGGTGAGAACTAATTTAGAAATGA 960  
| | | | |  
QY 968 ACAAGCAGATTAATGATGACAGTATGCTCTATAGACTCACTGCTGAATTAATAA 1027  
| | | | |  
Db 961 ACAAGCAGATTAATGATGACAGTATGCTCTATAGACTCACTGCTGAATTAATAA 1020  
| | | | |  
QY 1028 TGAAGTATTTTAATATGAAGATATGACACAGATATGATGATTTTGAAGAGCT 1087  
| | | | |  
Db 1021 TGAAGTATTTTAATATGAAGATATGACACAGATATGATGATTTTGAAGAGCT 1080  
| | | | |  
QY 1088 ATAGAGCTGCTTATGAAAAGTAACTCTATCTGCTATGCTGAACCTTTGTGCAAGTG 1147  
| | | | |  
Db 1081 ATAGAGCTGCTTATGAAAAGTAACTCTATCTGCTATGCTGAACCTTTGTGCAAGTG 1140  
| | | | |  
QY 1148 CTATTTTCACTGTGCTTTTAACTAATATGATGCTGCGCAATGAGGAAATGTGCGAG 1207  
| | | | |  
Db 1141 CTATTTTCACTGTGCTTTTAACTAATATGATGCTGCGCAATGAGGAAATGTGCGAG 1200  
| | | | |  
QY 1208 GTAACCTTACTGTGAGATCTAGTATGATGATGATGATGATGATGATGATGATGAT 1267  
| | | | |  
Db 1201 GTAACCTTACTGTGAGATCTAGTATGATGATGATGATGATGATGATGATGATGAT 1260  
| | | | |

```
Oy 1268 CCTGAACCTTTCTGGGAACTGTATATAGAGAGACTAGACAGCACTCATATATGAAACA 1327
    |||
Db 1261 CCTGAACCTTTCTGGGAACTGTATATAGAGAGACTAGACAGCACTCATATATGAAACA 1320
Oy 1338 CCTGTCTTACTCTACTCAAGAGTACACCCAAATTAAAGACAAAGTATGCTCCCC 1387
    |||
Db 1331 CCTGTCTTACTCTACTCAAGAGTACACCCAAATTAAAGACAAAGTATGCTCCCC 1380
Oy 1388 TTTCATATACACCAAGAGTACAGTACCGCTTGTATATATGATGATTTGATATCATTTG 1447
    |||
Db 1381 TTTCATATACACCAAGAGTACAGTACCGCTTGTATATATGATGATTTGATATCATTTG 1440
Oy 1448 AGGCGCAGAAAGTCTTGTAGTGAATATCTTTGAAAGTCCCTGACAGAAAGAAAGTGCCA 1507
    |||
Db 1441 AGGCGCAGAAAGTCTTGTAGTGAATATCTTTGAAAGTCCCTGACAGAAAGAAAGTGCCA 1500
Oy 1508 TTGTAGAGAGTATGAGTACAGGAAAGCAATATGTAGAGCTATTTTCTGCTCTATG 1567
    |||
Db 1501 TTGTAGAGAGTATGAGTACAGGAAAGCAATATGTAGAGCTATTTTCTGCTCTATG 1560
Oy 1568 AGCCCTCAAAAGGAGTATGATCTTGTCTGCTCAAAATATATACAGATGAGCCGAGAA 1627
    |||
Db 1561 AGCCCTCAAAAGGAGTATGATCTTGTCTGCTCAAAATATATACAGATGAGCCGAGAA 1620
Oy 1628 GCTTCGAGAGGCACTGGGAGTGTACCTCAGAGATGCTGCTCTTCCATATATCTATTT 1687
    |||
Db 1621 GCTTCGAGAGGCACTGGGAGTGTACCTCAGAGATGCTGCTCTTCCATATATCTATTT 1680
Oy 1688 ATTACACCTCTTATATGAAACATCATAGTCTTCACTGAGAGAGTATGAGTGGCAA 1747
    |||
Db 1681 ATTACACCTCTTATATGAAACATCATAGTCTTCACTGAGAGAGTATGAGTGGCAA 1740
Oy 1748 AATTAGCTGAGCTTATATGATATGCAATCTTCCAAATGCCAATGATATATGACCCAGTAG 1807
    |||
Db 1741 AATTAGCTGAGCTTATATGATATGCAATCTTCCAAATGCCAATGATATGACCCAGTAG 1800
Oy 1808 GGGAAAGAGAGTCAAGCTTTTCAGAGAGAGAAAGCAAAAGAGTACAAATGAGCAAGCCA 1867
    |||
Db 1801 GGGAAAGAGAGTCAAGCTTTTCAGAGAGAGAAAGCAAAAGAGTACAAATGAGCAAGCCA 1860
Oy 1868 TTTTGAAGAGACCCCAAGTCAATCTATATGATGAAAGTACTTCAATGTTAGATGATTA 1927
    |||
Db 1861 TTTTGAAGAGACCCCAAGTCAATCTATATGATGAAAGTACTTCAATGTTAGATGATTA 1920
Oy 1928 CTGAAGAGACTATTTCTGTGTGCATGAGAGAGTGTCAAAACAGAACTTATATTTCA 1987
    |||
Db 1921 CTGAAGAGACTATTTCTGTGTGCATGAGAGAGTGTCAAAACAGAACTTATATTTCA 1980
Oy 1988 TTGCACACAGATTTGCAACAGTGTGTGATGACAGATGAATCATTTGCTTGATCAGGTA 2047
    |||
Db 1981 TTGCACACAGATTTGCAACAGTGTGTGATGACAGATGAATCATTTGCTTGATCAGGTA 2040
Oy 2048 AGGTAGCCGAAAGTGTATACCAATGCTGTTGCTTCAACCTCATATGATCTATTCAG 2107
    |||
Db 2041 AGGTAGCCGAAAGTGTATACCAATGCTGTTGCTTCAACCTCATATGATCTATTCAG 2100
Oy 2108 AAATGTGCTATACAGAGACCGGTGTGCACAAACATGATTAACCCCAATGGAGCAA 2167
    |||
Db 2101 AAATGTGCTATACAGAGACCGGTGTGCACAAACATGATTAACCCCAATGGAGCAA 2160
Oy 2168 AGAAAGAAATATATCCAAAGAGAGAGAAAGAAAGTACAAAGAAAGAAATGTCATTA 2227
    |||
Db 2161 AGAAAGAAATATATCCAAAGAGAGAGAGAAAGAAAGTACAAAGAAAGAAATGTCATTA 2220
Oy 2228 GTGTGAAAAGCTGTGAAAAGTGTGTGTGTAAGTACATATAGACATTTCTTTTGTGT 2286
    |||
Db 2221 GTGTGAAAAGCTGTGAAAAGTGTGTGTGTAAGTACATATAGACATTTCTTTTGTGT 2278
```

RESULT 3  
US-09-212-247C-3  
; Sequence 3, Application US/09212247C

```
Patent No. 6391603
GENERAL INFORMATION:
APPLICANT: POMPEJUS, Markus; SUELBARGER, Harald; JOEFFKEN, Hans
            Wolfgang; DOVAL, Jose Luis Revuelta; JIMENEZ, Alberto;
            and GARCIA, Maria Angeles Santos
TITLE OF INVENTION: Genes of purine biosynthesis from Ashbya Gossypii
                    and the use thereof in microbial riboflavin
                    synthesis
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Kell & Weinkauff
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: IBM AT-compatible, Pentium processor
OPERATING SYSTEM: Windows 95
SOFTWARE: Mordperfect version 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/212,247C
FILING DATE: 16-Dec-1998
CLASSIFICATION: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5369 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..54
FEATURE:
NAME/KEY: CDS
LOCATION: 55..1482
FEATURE:
NAME/KEY: CDS
LOCATION: 1767..3299
FEATURE:
NAME/KEY: CDS
LOCATION: 3588..4703
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 4704..5369
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-212-247C-3

Query Match      15.3%; Score 358.6; DB 4; Length 5369;
Best Local Similarity 54.1%; Pred. No. 2.6e-94;
Matches 777; Conservative 0; Mismatches 649; Indels 9; Gaps 2;

Oy 703 AACCGATCTGGGTTTCACTGACAGACAGACGAGAGCTTATCTAAGCTATTGAC 762
    |||
Db 1 AAGCTGACCTTGCTGCTGACCTTATGTCGACAGAGTGACTTAACCCGCAATGGAT 60
Oy 763 AGAGAACAGAGGATATCAGTTTGTCTGAGTCTTGTGATTTAAATCTTCCATC 822
    |||
Db 61 CGTGTTTAAGGATATCTTATGTGCTGCTGCAATGTTTTCATATATCCGATT 120
Oy 823 ATGTTGAAGTATGCTTGTGAGTGTGTTTGTATTAACAATGCGGTGCCAGTTGCT 882
    |||
Db 121 ACATTTGAATATGATGATGATGAGCAATATGACATACCAAGTTGGTCTTCCCTGCT 180
Oy 883 TTGTAAACCTTGAACACTGTGATACATACACAGATTTCAAGTTGACAGACAGGTG 942
    |||
Db 181 GCTATTAACATTTGACATATCTTCTTACTTCATCTTTATCTTCAAGAACAGCGCGTGG 240
Oy 943 AGAAGTATTTAAGTAATGAATGAACAAAGCAGATATGATGACGATATGCTGTATA 1002
```

```
Db 241 CGCAGCGTTTGGCGTGAATGCGCAAGGCTGCAATTAAGGCGGTAGTGTGCATTTG 300
Qy 1003 GACTCAGCTGTAATTATGAAACTGTGAAGTATTTTAAATGAAGAATATGAACAG 1062
Db 301 GATTCCCTTAATTAATTTGAAAGCTGTAAAGTATTTCAATTAAGAGAAATACCTGGGAC 360
Qy 1063 AGATATGATGATTTTGAAGACGTATAGACCTGCTTATGAAGAAAGTACCTTATCTG 1122
Db 361 AAGTATCACACATCCTTGATGAAGTACCGGATTCAGATTAAGGTCTCCCAATCGCTG 420
Qy 1123 GCTATGCGAATTTGTGCAAGTCTATTTTCACTGCTGTTTACAGCTATATGTTG 1182
Db 421 GCGTTTGAACACCGGCAAGACCTAATTTTACACTGCACTGACGCAATGATGAT 480
Qy 1183 CTCGCCAGTCAAGGAAATGTGGCAGTACCCCTTACTGTGAGATCTGATATGTTGAT 1242
Db 481 ATGGCTGTAATGTTGTTATGCAAGGCTCTCTTAACATGCGGAGATCTGTGTTAAT 540
Qy 1243 GGAATGCTTTTCACTTCAATTAACCTGAACTTTCTGGAACTGATATAGAGACT 1302
Db 541 CAATGCTATTCAGACTCTCGGTGCACTAACTCTTGATAGGCTTACCGTATCTC 600
Qy 1303 AGACAGACTCTATGATATGATGAACCTTTGTTACTTACTCAAGTGAACCAATTT 1362
Db 601 AAGAGTCTGATATGATATGAAATCTTTATTAACGCAAAAATCAGGTCAATTT 660
Qy 1363 AAAGCAAAAGTATGAGTATCTCCCTTCAAGATCAACACAGACGATACGGTGCCTT 1422
Db 661 AAGAACTCCCAATATGCCAGAACTTACCAATTAACAAACCGTTGATATTTGCTTT 717
Qy 1423 GATATGTCATTTTGAATATGATGAGGCGCAAGAACTCTTAATGAAATATCTTTGAA 1482
Db 718 GAAATGTTAGCTTTGGCTATGACCCGAGGCGGTATATTAACAATGTTTGGTTACC 777
Qy 1483 GTCCCTGAGAGAAAGAGTGGCCATTTAGAGGATAGTGGCTCAGGAAAGACAAATA 1542
Db 778 ATCCAGCTGGAATGAAACATGCAATAGTACGCCATCGGCTCGGGGAATCCACCAAT 837
Qy 1543 GTGAGGCTAATTTTGGCTTATGAGGCTCAAAAAGGTAGATTTATCTTGCTGTCAA 1602
Db 838 TTGAAGCTCGATTTATGATTTCTAAGAGCCGAGCAAGGTCTGATTCCTAGTGGGCA 897
Qy 1603 AATATACAAATGATGAGCTGAAAGCCCTTGAGAGGAGTGAAGTGTACTCTAGAT 1662
Db 898 GATATCGCGATTTAGACTTGTCTTTCTTACGAAAGCTATCGGTGTGCCCAAGAT 957
Qy 1663 GCTGCTCTTCTCATATATCTATTTATTAACAACCTTTATATGAAACATAGTCTTCA 1722
Db 958 ACTCCTCTTCAATGACACATCTGGAGATGTTAATTCGGCAATATCAGTTCCTCT 1017
Qy 1723 CCGTGAAGATGATGAGTGGCAAAATTTAGCTGACTTCAATGATGCAATTTCTTGATG 1782
Db 1018 GACGATGATTTCTCAGGCGCATTAAGAAAAGCTCAACTCAGAACTATCTCAAGACCTA 1077
Qy 1783 CCAGATGATATGACACCAAGTAGAGGAGAGAGAGCTCAAGCTTTAGAGAGAAAG 1842
Db 1078 CCAAGAGCGCTTCCACCGTTGTAGGAGGAGCGGCTTTGATGATCAGGAGGAGGAAA 1137
Qy 1843 CAAAGAGTGAATGCAAGGAGCAATTTTGAAGACCCCGCAATATCTCTATGATGA 1902
Db 1138 CAAAGGCTGTATGCTGCTGTGTGTTTGAAGAGCGCTCGGATGTTTTCAGAGAG 1197
Qy 1903 GCTACTCATGCTTATGATGATTTAGTGAAGAGACTATTTCTGTGTCAT-----GAG 1956
Db 1198 GCTCAAGTGTCTGGAATACACACAGAGAGGACTTTTGCAACCAATTCACAGAAC 1257
Qy 1957 GATGTGCTCAACACAGAACTTTCTATTTTCAATGACACAGATTTGTCAACAGTGGTAT 2016
Db 1258 TTTTCTTCAATTCAGAGAGAGGTTTACGTTGCCATACATCGCGACATAGCTGAT 1317
Qy 2017 GCAATGAATCATTTGCTTGTGATCAGGATAGGTAGCCGAACGTGTTACCCACATG 2076
```

```
Db 1318 GCAGATTAAGTATGTTGTTGAAAGGTTCTGTCCGCGAAGAGGCGACACAGACTCG 1377
Qy 2077 TTGCTGCTAAACCTCATAGTATCTATTTCAAGAAATGTGGATACAGAGAGCC 2131
Db 1378 CTGTAGCGTACAGAGATCCCTATACCGGGGTGTGTGGATATTCAGAAAAACC 1432

RESULT 4
US-09-134-001C-1893
; Sequence 1893, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1893
; LENGTH: 1749
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1893

Query Match 8.1%; Score 188.8; DB 4; Length 1749;
Best Local Similarity 49.2%; Pred. No. 4.9e-45;
Matches 524; Conservative 0; Mismatches 537; Indels 3; Gaps 1;

Qy 1026 TGTGAATATTTAATATGAAGATATGAAGCAGAGATATGATGATTTTGAAGAC 1085
Db 648 TATTAAGATTGCTATGTAAGCAATGAGCTTAATAATTTGATTAACATAACAGAA 707
Qy 1086 GTATGAGACTGCTTCAATGAAAAGTACCTTACTCTGCTATGCTGAATTTGTCTAAG 1145
Db 708 TTTTTCACAGAGCTTCCACATACAGATGAGAGCATATTTCTTGTGCTATTTAA 767
Qy 1146 TGTATTTTCAAGTGTGCTTTTAAAGCTATATGATGCTGCGCAGTGAAGAAATGTC 1205
Db 768 TACTGTATCAGATTTAGGCCCAATTAATGATGAGTGGGTGCTCATATTTGGCAATTAC 827
Qy 1206 AGGTACCTTACTGTGAGATCTAGTATGATGATGATGATGATGATGATGATGATGAT 1265
Db 828 AGATGATTAATCTGTGGAACCTTACGAGCATTTGTGCTTATCTAGAACATATTTGG 887
Qy 1266 ACCCTGAACCTTCTGGAACCTGATATGAGAGACTAGACAGCACTCATATGATGA 1325
Db 888 ACACATTAAGAGCTATGATCTTCACTTACTATCACTTACAAAGTTTGCATATGGA 947
Qy 1326 CACCTGTTTACTTACTTCAAGGTAGACACCAATTAAGAAAGCAATGATGCTATCC 1385
Db 948 CAGAGTATTTCAAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1007
Qy 1386 CTTTCAATACACACAGACAGTACCGTGGCTTTGATATGATGATGATGATGATGAT 1445
Db 1008 AATTAATAGTAAAGGTCAAATTTGATTAACATGATGATGATGATGATGATGATGAT 1067
Qy 1446 TGAAGGCGAAGAGTCTTATGATGATATCTTTGAAGTCCCTGAGAGAAAGTGGC 1505
Db 1068 TGA---AAAAGATATTAACAGATATTAATTTAACAATTAACAAAGCGAACTGAGC 1124
Qy 1506 CATTTGAGAGTATGAGTCAAGGAAAGACAAATGATGATGATGATGATGATGATGAT 1565
Db 1125 ATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1184
Qy 1566 TGAAGCTCAAAAGGATGATTTATCTGTGATCAAAATATACAGATGATGATGATGAT 1625
Db 1185 TGAATGATCAAGGTGAATTAATCTTATGATCAATCAATTAATTAATTAATTAATTA 1244
```

QY 1626 AACCTTCGAGGAGTGGAGTGTACTCTCAGATGCTGCTCTTCATTAATCTAT 1685  
DB 1245 TAGTTAAGAAATCAAAATAGCTTAGTACACAGATATATCTTTTCTGATACGCT 1304  
QY 1686 TTATTACAACCTCTATATGAAACATCAGTCTTCACTGAGGAAGTATAGCAGTGC 1745  
DB 1305 TAAAGAGATATATTTTGTGGTAGGCTGATGCGACTGATGATGAAGCTGAC 1364  
QY 1746 AAAATTAGGTGACTTTCATGATCAATCTTGAAATGCCATGATATGACACCCAGT 1805  
DB 1385 AAAAATGGCGAATGCCCATGATTTTATTTCAAAATTTACCGAATGATATGATCGAAGT 1424  
QY 1806 AGGGAGACGAGACTCAAGCTTTTCAGAGAGAAAAGCAAGATGACATTTGCAAGAC 1865  
DB 1425 AGGAGACGAGAGAGTTAATATCTGCTGAGCAAAAACAAAGTTGTCAATTTGACGAT 1484  
QY 1866 CATTTGAAGAGCCCCCAGTCATCTCTATGATGAAGTACTTATCTTATGATTCGAT 1925  
DB 1485 CTTTAAATATATCTCTCTGTTTATATATGATGAGCAACAGTGCATTTGATTTAGA 1544  
QY 1926 TACGTAGAGACTATTTCTTGTGCGATGAGAGATGCTGCAACACAGAACTTCTATTTT 1985  
DB 1545 GGTGAGCTATATTTCAAGAACCTGATGTTTAACTAGATAGAACAACTTAAT 1604  
QY 1986 CATTCACACAGATTGTCAACAGAGTGTGATGACAGATGAAATCATTTGCTTGATCAGG 2045  
DB 1605 TGTTCACATGCTCTATCTATCACTACTCATGACAGATAGATAGTTGATGGAATAATG 1664  
QY 2046 TAAAGTAGCCGACGTTGATCCACCATGTTGCTTCTGCTAAC 2089  
DB 1665 ACGAATTTGTGAGACTGGACACACACCAATTAATTAATTAAC 1708

## RESULT 5

US-08-961-527-63  
; Sequence 63, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunesh  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 63:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7760 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

US-08-961-527-63

Query Match 7.7%; Score 180.2; DB 4; Length 7760;  
Best Local Similarity 54.0%; Pred. No. 3,7e-42;  
Matches 368; Conservative 0; Mismatches 313; Indels 0; Gaps 0;

QY 1394 TCACACACAGACAGCTACCGTGGCTTTGATTAATGTCATTTTGAATACATTAGGCCC 1453  
DB 6480 TCACTAAGTTSCAAGAAAGTTGAAATCAAGTCATATCGTTTTCATCTACTTGCCTGATA 6539  
QY 1454 AGAAAGTCTTAATGGAATATCTTTGAAGTCCCTGCAGGAAAGAAACTGGCCATTGAG 1513  
DB 6540 AACCTATTTTAAAGATGTCAGATTTCTGCTTAAAGGCCAATGACACAGCTGTG 6599  
QY 1514 GAGGTAGTGGTCAAGGAAAGCAATAGAGGCTATTATTTGCTTATAGGCTC 1573  
DB 6600 GCGGACAGGTTCCAGGAAAGACACTATATGAACTCATATCGCTTTATATAGTGG 6659  
QY 1574 AAAAGGTAGCATTTATCTTGTGTCAAATATACAAAGATGAGCTGGAAAGCCTTC 1633  
DB 6660 ATGCTGGTGTATTTATTTGATGATGTAAGACATTCGTGCTATGACTTATAGTCTTA 6719  
QY 1634 GAGGCGAGTGGAGTGTACTCAGATGCTGTCTCTTCATATATCTATTTATACA 1693  
DB 6720 GAACCAAGTGGGAATGTATGCAAGATGCTTGTGTTAGCGGAACGATTAAGACA 6779  
QY 1694 ACCTTTATGAAACATCAGTCTTCACTGAGAAAGTATGACAGTGGCAAAATTAG 1753  
DB 6780 ATATCCATTTGTTGTGTGCGAATGCTATGCAAGAAATGTTGAGAGTACAGCAAAAGCA 6839  
QY 1754 CTGACTTATGATGCAATTTCTTGAATGCCAATGATATGACACCCAGTAGGGGAC 1813  
DB 6840 CCCACATTCAGACTATATTCGAAAGTTGCTGATTAAGATGATATCTTATATGATG 6899  
QY 1814 GAGAGCTTAAGCTTTCAGAGAGAGAAAGCAAGATGATGCAAGAGCCATTTTGA 1873  
DB 6900 ACCAGAGATCTTTTCAACAGGAGCAAGCAATGATTTCAATGCTGCAACCCGATGA 6959  
QY 1874 AGGACCCCCAGTACTACTATGATGAGAGCTACTTCACTGTTAGTTGATCTGAG 1933  
DB 6960 CAGATCCAGAAATCTCATTTCTCGATGAGCAACTTCAAGATGATACGCTGACAGAA 7019  
QY 1934 AGATATTTCTTGTGCGATGAGAGATGTGTCAACACAGAACTTATTTTATTTGAC 1993  
DB 7020 GCAAGATTCAGATGCCATGAGAGGTGTTGACAGAGTAACTAGTTTCTCATTTGCC 7079  
QY 1994 ACAGATTTCAACAGTGTATGACAGTGAATCATTTGCTTGTGATCAGGTTAAGTAG 2053  
DB 7080 ACCGCTTGAACCAATCTCATGACATGCAATTAATTTGCTTAAAGATGAGAAAGTCA 7139  
QY 2054 CGAAGCGTGTACCCACCATG 2074  
DB 7140 TTGAACGTGTACCCACCATG 7160

## RESULT 6

US-08-772-270A-6  
; Sequence 6, Application US/08772270A  
; Patent No. 6019984  
; GENERAL INFORMATION:  
; APPLICANT: MacInnes, Janet  
; APPLICANT: Ricciardi, Paul  
; APPLICANT: Mallard, Bonnie  
; APPLICANT: Rosendal, Soren  
; TITLE OF INVENTION: NOVEL BACTERIAL PREPARATIONS, METHOD FOR  
; TITLE OF INVENTION: PRODUCING SAME, AND THEIR USE AS VACCINES  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bereskin & Parr  
; STREET: 40 King Street West  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada

ZIP: MSH 3V2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/772,270A  
FILING DATE: December 23, 1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Micheline  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 6580-81  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5120 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: Actinobacillus pleuropneumoniae  
US-08-772-270A-6

Query Match 7.2% Score 170; DB 3; Length 5120;  
Best Local Similarity 55.1%; Pred. No. 2.8e-39;

Matches 332; Conservative 0; Mismatches 270; Indels 0; Gaps 0;

QY 1502 TGGCCATTGTAGAGTGTGGGTGCGAGGAAAGACATAGTGGCTATTATTTGCGT 1561  
DB 2084 TCGGATGTGAGAGCTTCAGGCTCAGGAGAGAGACCTTACGAAATTAATTCAGCTT 2143  
QY 1562 TCTATGACCTCAAAAGGGTGTGACATTATCTTCTGTCAAAATATACAGATGAGCC 1621  
DB 2144 TTTTATATCCGAAAACCGTACGATTAATGATGGGCAATGATTTAGCATTTGCGGATC 2203  
QY 1622 TGGAAAGCCTTGAGGGGCGAGTGGGAGTGTACCTCAGAGATGCTGCTCTTCATATA 1681  
DB 2204 CGAACTGCTAGTGTCTAAGTCGGGGGTGTATTAAGATTAACGTAATTAATTCGTA 2263  
QY 1682 CTATTTATTACACCTCTTATATGGAACATCAGTGTCTCACTGAGGAGTGTATGAG 1741  
DB 2264 GTATTCGAGATATATTCCTTAGGGGATCCGGGTATGCCAATGAAAAATTTCCATG 2323  
QY 1742 TGGCAAAATAGCTGAGCTTCATATGATGCAATTTTGAATGCCAATGATATGACACC 1801  
DB 2324 CGGCAAAATAGCCGGCGCACATGAAATTTATCTGAATTCGTGAGGATATTAACACGA 2383  
QY 1802 AAGTAGGGGAACGAGACTCAAGCTTTCAGAGAGAAAAAGCAAGAGTATGCAATTTGCA 1861  
DB 2384 TTGTGTGAGACAGAGTGGGGGTATCTGGCGGCAACGCAACGATTTGCGATTTGCAC 2443  
QY 1862 GAGCATTTTGAGAGACCCCGCAGTCACTATCTATGATGAGTACTTCACTGTTAGATT 1921  
DB 2444 GCGCTTTGTGATTAACCCGAAATCTTAATTTTGTGATGAAGCAACAGCGCATTTAGATT 2503  
QY 1922 CGATTAAGAGACTATTTCTTGTGTCATGAAGAGATGTGTCAACACAGAACTTCTA 1981  
DB 2504 ATGAATCCGAGCATATCATCATGCGCAATATGACACAGATTTTAAAGGAGAAACGGTAA 2563  
QY 1982 TTTTCATTGACACAGATGTGCAACAGTGTGATGCAAGATGAATCATTTGTCTTGATC 2041  
DB 2564 TTATCATTTGACACCGTTTATCTACGGTAAATAATGCGAACCGTATTAATTTGTATGAAA 2623  
QY 2042 AGGTTAAGTACCGAAGTGTACCAACATGTTGCTTGTGTAACCTCATAGTATCT 2101  
DB 2624 AAGTGCAGATTTGTGAACAGTATGACATTAAGAGCTGCTTGTGATTCACAAACGGCTTAT 2683  
QY 2102 AT 2103

DB 2684 AT 2685

RESULT 7

US-08-488-706-1  
Sequence 1, Application US/08488706

Patent No. 5994525

GENERAL INFORMATION:

APPLICANT: Kamp, Elbarte M.

APPLICANT: Smits, Marinus A.

TITLE OF INVENTION: Recombinant Vaccine For Prevention and/or

TITLE OF INVENTION: Treatment of Pleuropneumonia Infections

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Abelman, Frayne & Schwab

STREET: 708 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10017

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

COMPUTER: IBM PC or compatible

OPERATING SYSTEM: DOS 3.31

SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,706

FILING DATE: 09-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/722,971

FILING DATE: 28-JUNE-1991

ATTORNEY/AGENT INFORMATION:

NAME: Cinamon, Jay S.

REGISTRATION NUMBER: 24,156

REFERENCE/DOCKET NUMBER: 201,875

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-949-9022

TELEFAX: 212-949-9190

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 8370

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-488-706-1

Query Match 7.2% Score 170; DB 2; Length 8370;  
Best Local Similarity 55.1%; Pred. No. 3.7e-39;

Matches 332; Conservative 0; Mismatches 270; Indels 0; Gaps 0;

QY 1502 TGGCCATTGTAGAGTGTGGGTGCGGAAAGCACATAGTGGCTATTATTTGCGT 1561  
DB 5334 TCGGATTCGAGAGAGCTTCAGGCTCAGGAGAGACCTTAACGAAATTAATTCACGTT 5393  
QY 1562 TCTATGAGCTCAAAAGGGTGTGACATTATCTGCTGCAAAATATACAGATGTGAGCC 1621  
DB 5394 TTTTATATCCGAAAACGGTACAGTATTAATGATGGGCGATGATTTAGCATTTGCGGATC 5453  
QY 1622 TGGAAAGCCTTGAGAGGCGAGTGGAGTGTACCTCAGATGCTGTCTTCCATATA 1681  
DB 5454 CGAACTGCTAGCTGTGCAAGTGGGGTGTATTAAGATTAAGTATTAATTTAAATGTA 5513  
QY 1682 CTATTTATTACACCTCTTATATGGAACATCAGTGTCTTCACTGAGGAGTGTATGAG 1741  
DB 5514 GTATTCGAGATTAATTTGCTTACCGGATCCGGGTATGCAAAAGTAAATTTGTCATG 5573  
QY 1742 TGGCAAAATAGCTGAGCTTCATGATGCAATTTCTGGAATGCCAATGATATGACACC 1801  
DB 5574 CGGCAAAATAGCCGGCGCACATGAATTTATTTCTGAATTCGTGAGGATATTAACACGA 5633  
QY 1802 AAGTAGGGGAACGAGACTCAAGCTTTCAGAGAGAAAAAGCAAGATGCAATTTGCA 1861





REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3576 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-976-259-79

Query Match 7.1%; Score 165.4; DB 4; Length 3576;  
Best Local Similarity 53.7%; Pred. No. 5e-38;  
Matches 340; Conservative 1; Mismatches 292; Indels 0; Gaps 0;

QY 1492 GGAAGAAAGTGGCCATTGTAGAGGTAGGCTCAGAGGAAACACAAATAGTAGGCTA 1551  
DB 3564 GGGAGGTTATGGTATGTCGACGTTCTGGTTCAGAGAAACACATTAATACTAAATTA 3505  
QY 1552 TTATTTGCTTCTATGAGCCTCAAAAGGGTAGCATTTATCTTGCTGGTCAAAATATACAA 1611  
DB 3504 ATTCAAGTTTTTATCTCCGAAATGCGCAGGTCTTAATTGATGACATGATCTTGGC 3445  
QY 1612 GATGAGCCTGGAAGCCTTCGAGGGCAGTGGAGTGTACTCAGATGCTGCTCC 1671  
DB 3444 TTGGCCGATCTTAAGTGTACGTGTCAGGTGGGGGTGTGTTGTCAGGACATGTGCTG 3385  
QY 1672 TTCATATATCTATTTATTAACACTCTTATATGAAACATCACTGCTTCACTGAGGAA 1731  
DB 3384 CTATATGCGCATATATATGATTAATTTCACTGCTATCTCGCATGTCCTCGGAAAAA 3325  
QY 1732 GTATGAGAGTGGCAAAATGATGCTGAGCTTATGATGCAATTTCTGAAATGCCAGTGA 1791  
DB 3324 GTTATTTATGAGGCAATTTASCAAGTGTCTATGATTTTCTGATTTGCGTAGGGG 3285  
QY 1792 TATGACACCCAGTAGGGAACGAGACTCAAGCTTTGAGGAGGAAAGCAAGAGTA 1851  
DB 3264 TATTAACACATGTGCGGGAAACAGGGGCAAGATTATCCGAGGTCAACGTCAACGATC 3205  
QY 1852 GCATTTGCAAGACCATTTTGAAGACCCCCAGTCACTCTATGATGAAGTACTTCA 1911  
DB 3204 GCATTTGCAAGGGCGTGGGAAACACCTAAATTAATCTCATCTTTGATGAAGCAACAGT 3145  
QY 1912 TCGTATGATTCGATTAATGAGAGACTATCTTGTCGATGAAGATGTCGAAACAC 1971  
DB 3144 GCTTGAATTTATGATGAGAGCATGTCATCAATGCGCAATATGCAAAATATGTAAAGGC 3085  
QY 1972 AGAATCTTATTTTCAATGCAACAGATTGTCAACAGTGTGATGAGATGAATCAT 2031  
DB 3084 AGAAGGTTATATCATTTGTCATGCTGTCTACAGTAAATAATGAGACCGGATTAAT 3025  
QY 2032 GTCTTGATCAGGGTAGTACCGCAAGCTGTGTAACCAACATGCTTTGCTTCAACCT 2091  
DB 3024 GTATGAGAAAGGAAATTTGTGAACAGGGTAAACATTAAGAGCTGCTTTCGAACCG 2965  
QY 2092 CATGATCTATTGAGAAATGTCATACACAG 2124  
DB 2964 GAAAGTTATACAGTTACTTATATCAGTTACAG 2932

RESULT 10  
US-08-961-527-188  
Sequence 188, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue

CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 188:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6492 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-188

Query Match 6.6%; Score 154.8; DB 4; Length 6492;  
Best Local Similarity 50.0%; Pred. No. 8.7e-35;  
Matches 387; Conservative 0; Mismatches 387; Indels 0; Gaps 0;

QY 1301 CTAGACAGACTCATATATGATGAACACTTGTACTACTCAAGTAGACACCCAAA 1360  
DB 2876 CTAAAGAGGCTTGATGAAATATTTATTTATCGCCAAATAGAAAGATTAAAGAAA 2935  
QY 1361 TTAAAGCAAGGATGAGCATCTCCCTTCAGATCAACACAGACAGCTACCGTGCC 1420  
DB 2936 TTCAAAATCAAGTTTACAAAGGAGTACTATAGCTTAAATAATTTGATATGATC 2995  
QY 1421 TTGATATGTCATTTTGAATATCATTTAGGGCCAGAAAGTCTTAGTGAATATCTTTG 1480  
DB 2996 TAAAGATGTGAGTTTGCTTACATTAAGACCAAAAGTTTAATGTGATGATTTTA 3055  
QY 1481 AAGTCCCTGAGAAAGAAAGTGGCAATTTAGAGATGTGGTCAAGGAAAGACAA 1540  
DB 3056 AAGCTAAGCAGGAGAGTCACTGCTTTGTAGTGCAGAGTGGCTCGGTAACAACTA 3115  
QY 1541 TACTGAGGCTATTTTCTGCTTATGAGCCTCAAAAGGTACATTTATCTTGCTGCTC 1600  
DB 3116 TCTTGAACCTTATATCAAGCTTTATGATTAATGACAAGGACAAATCTTAATGATGCA 3175  
QY 1601 AAATATACAAAGTGTAGCCTGGAAGCCTTGCAGGCAAGTGGAGTGTACTCAGG 1660  
DB 3176 AATATTAAGAAATATCAACGAATCCCTTTTGATTAAGGTGTCTATTTGTTCCAG 3235  
QY 1661 ATGCTGCTCTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1720  
DB 3236 ATGTGCTCTCTTAATCAAAAGGCTTATGAAATTAATTAATTAATTAATTAATTAAT 3295  
QY 1721 CACTGAGAGATGTATGACGTGGCAAAATTAAGTGTGATCTATGATGCAATCTTCGAA 1780  
DB 3296 GTGACGAAGGTTAAAGAGCAGCAAACTTGCAATTTGACAGATTTTATTAAGAAAAA 3355  
QY 1781 TGCCATGAGATATGACACCAAGTAGGGGAAAGAGACTCAAGCTTTGAGAGAGAAA 1840  
DB 3356 TGATTAAGTTTCAATACATGATTTGTGAAAGAGAGCTGATGATATCAGAGAGGAAA 3415  
QY 1841 AGCAAGAGTAGCATGCAAGAGCAATTTTGAAGACCCCGGTCATATCTATGATG 1900

Db 3416 GACAAAGATTATCAATAGCAGAGCCTTCTTAAAGATGCCGATATGATCTTAGATG 3475  
Oy 1901 AAGTACCTTCATCGTATGATTCGATTAAGAGACTATTCCTGTGTCATGAGATG 1960  
Db 3476 AGATACACGCAAGCCCTTGATGTTAAACAGAGAAAGATTAAGAGCTTTAAATAAT 3535  
Oy 1961 TGGTCAACACAGAACTTCTATTTTCATTTGCACACAGATTGTCAACAGTGTGATGAG 2020  
Db 3536 TAGTTAAAGATTAACCTGTTGATTCATTTTCATAGATGAATCCATAGAAATGCG 3595  
Oy 2021 ATGAATCATCTTCTTGATCAGGGTAAGTACCGAACTGTACCAACATG 2074  
Db 3596 ACAAGATAGTGTCTTCAAAACGAGAGATGAAGAAAGCAAGTAAAGATGAMG 3649

RESULT 11  
US-08-961-527-23/c  
Sequence 23, Application US/08961527  
Patent No. 6420135

## GENERAL INFORMATION:

APPLICANT: Charles Kunesh  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

TELECOMMUNICATION INFORMATION:

REFERENCE/DOCKET NUMBER: PB340P1

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 7147 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-23

Query Match 6.3%; Score 147; DB 4; Length 7147;

Best Local Similarity 53.9%; Pred. No. 1,7e-32;

Matches 325; Conservative 0; Mismatches 275; Indels 3; Gaps 1;

Oy 1482 AGTCCCTGAGAGAAAGAGTGGCCATTTAGAGAGTAGGTGACAGGAAAGACAAT 1541  
Db 798 AGCTATGCTGTCGCAAAAGGTGCCATTTTGGCCGACTGGGGCTGTAACAACAT 739  
Oy 1542 AGTGAAGCTATTATTCGCTTATGAGCCTCAAAAGGATAGCATTTATCTGCTGTC 1601  
Db 738 TGTTCATCTTTTGATGAATTTGATGATGATTAAGGAAGATTCGATTCATGATG 679  
Oy 1602 AATATATCAAGATGTGAGCTGGAAGCCTTGGAGGGAGAGTGGAGTGTACCTCAGA 1661  
Db 678 GGATACCAAGGCTATGACCGCTTCAAGAGTGAATGATGCTTTTCAATGCTCTTG 619

Oy 1662 TGTCTGCTCTTCATTAATATCTATTTATTAACAACCTCTATATAGAAATCAAGTCTTC 1721  
Db 618 TACTGCTCTTTTGAAGAACTATTCGAGAACATCTCATATATCAATATAGGATTAG 559  
Oy 1722 ACCTGAGGAAGTGTATGACGTGCGAAATTAAGCTGAGCTTCATGATGCATTTCTG 1781  
Db 558 TGAATGAACGATGTATGAACTAGTAAGCTGTGGGAATTCACCATTTATATGACCTT 499  
Oy 1782 GCCACATGATATGACACCCCAAGTAGGGGAACGAGACCTCAAGCTTTACAGAGAGAAA 1841  
Db 498 GCCAGATGCTATATATACATCTTGATGACACCG---TACCTGTCTGTAAAGACAAA 442  
Oy 1842 GCAAAAGTAGAATTCGCAAGAGCCATTTTGAAGACCCCCAGTCACTATCTATGATGA 1901  
Db 441 ACACTATATGACTATGCTGCTGCTCTTTAAAGATGACACCGCTTTGATTTGATGA 382  
Oy 1902 AGCTACTTCATCTGTTAATTCGATTTACTGAAGACATATCTTGTGCTCAAGATG 1961  
Db 381 GGCACCTTCTTCTGTGTAACACAGCAGAGAGAAATGATCCAAAACCATGACCGTTT 322  
Oy 1962 GGTCAACACAGAACTTCTATTTTCATTTGCACACAGATTGTCAACAGTGTGATGAGA 2021  
Db 321 GATGAAGAGACGACATCTTGTTCATTTGCCACCGCTTGTCAACATCCGAATGAGA 262  
Oy 2022 TGAATCATCTTCTTGATCAGGGTAAGTAGCCGAACTGTATCCCAATGCTTGTCT 2081  
Db 261 CTGATCTTGTGTATGAAGATGAAATATCATGACGACCAAGCAATATGAGAACTGAT 202  
Oy 2082 TGC 2084  
Db 201 GGC 199

RESULT 12

US-08-461-823-1

Sequence 1, Application US/08461823

Patent No. 5593840

GENERAL INFORMATION:

APPLICANT: Bhatnagar, Satish K.

APPLICANT: George Jr., Albert L.

APPLICANT: Nazarenko, Irina

TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACID SEQUENCES

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: OncoPharm, Inc.

STREET: 200 Perry Parkway

CITY: Galthersburg

STATE: Maryland

COUNTRY: USA

ZIP: 20877

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/461,823

FILING DATE: 05-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/168,621

FILING DATE: 16-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/010,433

FILING DATE: 27-JAN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Karta, Glenn E.

REGISTRATION NUMBER: 30,649

REFERENCE/DOCKET NUMBER: PA-0012 CIP 2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301 527-2058

TELEFAX: 301 208-6997

INFORMATION FOR SEQ ID NO: 1:

```
SEQUENCE CHARACTERISTICS:
LENGTH: 2726 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-461-823-1

Query Match      6.1%; Score 143.8; DB 1; Length 2726;
Best Local Similarity 52.9%; Pred. No. 8.6e-32;
Matches 335; Conservative 0; Mismatches 292; Indels 6; Gaps 1;

QY 1458 AGTCCTTAGTGAATATCTTTGAAGTCCCTGAGAAAGAGTGGCATTGTAGAGG 1517
    |||
DB 1457 AGTCTTCAGGAGCTGAGCCCTGAGAGTGAAGAAGGGCCAGCGTGGCTGTGGGGAG 1716
    |||
QY 1518 TAGTGGGTGAGGAAAAGCAATAGTAGGCTATTTATTTGCTTCTATGAGCCTCAAAA 1577
    |||
DB 1717 CAGTGGCTGTGGGAAAGACAGAGTGTCCAGCTCTGGAGCGGTTCTACGACCCCTTGGC 1776
    |||
QY 1578 GGGTAGCATTTATGCTGTGTCAAAATATACAGATGTAGGCTGGAAAGCCTTGGAG 1637
    |||
DB 1777 AGGAAAAGTCTGCTTGAAGCAAAATAAAGCACTGATGTTCAGTGTCCGAGC 1836
    |||
QY 1638 GGCAGTGGAGTGTGTACTCAGATGTGTCTCTCCATATATCTATTTATTAACCT 1697
    |||
DB 1837 ACACCTGGGAGTGTGTCCAGAGGCCCATCTGTGTAAGTGAAGTGTGTGAGAACAT 1896
    |||
QY 1698 CTTATATGAAACATCAG-----TGCTTCACTGAGAGTGTATGCAATGSCAAAATT 1751
    |||
DB 1897 TGCCTATGAGAACAAAGCCGGGTGTGTCAAGAAAGATCGTGAAGGAGCAAAAGGA 1956
    |||
QY 1752 AGCTGACCTCATGATGCAATTTCTTGAATGCCACATGATATGACCCCAAGTAGGGA 1811
    |||
DB 1957 GGGCAACATACATCTCTTCATCGAGTCACTGCTAATTAATATGACCTAAAGTAGAGA 2016
    |||
QY 1812 ACGAGACTCAAGCTTTGAGAGAGAAAGCAAGTAGCAATTTCCAAAGGCCATTCTT 1871
    |||
DB 2017 CAAGGAACCTAGCTCTCTGTGTGCGCAAAACAAAGCATTCCTAGCTGTGCTGTGT 2076
    |||
QY 1872 GAAGGACCCCACTCATCTATCTATGATGAAGTACTTCATGCTTAATTCGATTACTGA 1931
    |||
DB 2077 TAGACACCTCATATTTTGTGTTGATGAAGCCAGCTCGATACAGAAAGTGA 2136
    |||
QY 1932 AGAAGCTATTTTGTGTGCGCATGAAGATGTGTCAACACAGAACTTCTATTTTCATTGC 1991
    |||
DB 2137 AAAGGTGTCCAAAGAGCCCTGGAACAAGCCAGAGGCCGACCTGTGCAATTTGATTGC 2196
    |||
QY 1992 ACACAGATTGTCAACAGTGTGTGATGAGATGAATCATTTGCTTGGATCAGGTTAAGT 2051
    |||
DB 2197 TCACCGCTGTCCACATCCGAATGCAAGTCTTAATAGTGTGTTTCAGAAATGCAAGT 2256
    |||
QY 2052 AGCCGAACGTGTACCCACCATGTGTTGCTTGC 2084
    |||
DB 2257 CAAGGACATGCGACGCAATCAGCAGCTGCTGGC 2289
    |||

RESULT 13
US-08-784-649A-1
; Sequence 1, Application US/08784649A
; Patent No. 5830697
; GENERAL INFORMATION:
; APPLICANT: Sikic, Branimir I
; APPLICANT: Chen, Gang
; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
; CYCLOSPORIN MODULATION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road
; CITY: Menlo Park
```

```
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,649A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: Reg.No. 5830697 36,677
REFERENCE/DOCKET NUMBER: 06037/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4264 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-784-649A-1

Query Match      6.1%; Score 143.8; DB 2; Length 4264;
Best Local Similarity 52.9%; Pred. No. 1.1e-31;
Matches 335; Conservative 0; Mismatches 292; Indels 6; Gaps 1;

QY 1458 AGTCCTTAGGATATCTTTGAAGTCCCTGAGAAAGAGTGGCATTGTAGAGG 1517
    |||
DB 3291 AGTCTTCAGGAGCTGAGCCCTGAGAGTGAAGAAGGCCAGACCTGTGTGGGGAG 3350
    |||
QY 1518 TAGTGGGTGAGGAAAAGCAATAGTAGGCTATTTATTTGCTTCTATGAGCCTCAAAA 1577
    |||
DB 3351 CAGTGGCTGTGGGAAAGGACAGTGTGTCCAGCTCTGAGCGGTTCTACGACCCCTTGGC 3410
    |||
QY 1578 GGGTAGATTATCTTGTGTGTCAAAATATACAGATGTAGAGCTTGAAGACCTTGGAG 1637
    |||
DB 3411 AGGAAAAGTCTCTTGAATGCAAAAGAAATTAAGCACTGAATGTTTCAGTGTCTCGAGC 3470
    |||
QY 1638 GGCAGTGGAGTGTGTACTCAGATGTGTCTCTCCATATATCTATTTATTAACCT 1697
    |||
DB 3471 ACACCTGGGAGTGTGTCCAGAGGCCCATCTGTGTAAGTGAAGTGTGTGAGAACAT 3530
    |||
QY 1698 CTTATATGAAACATCAG-----TGCTTCACTGAGAGTGTATGCAAGTGGCAAAATT 1751
    |||
DB 3531 TGCCTATGAGAGCAACAGCCGGGTGTGTCAAGAAAGATTTGTAGGSCAGCAAAAGGA 3590
    |||
QY 1752 AGCTGACCTCATGATGCAATTTCTGAAATGCCACATGTATGACCCCAAGTAGGGA 1811
    |||
DB 3591 GGGCAACATACATGCTTCATCGAGTCACTGCTTAATTAATATAGACTTAAGTAGAGA 3650
    |||
QY 1812 ACGAGACTCAAGCTTTGAGAGAGAAAGCAAGTAGCAATTTGCAAGACCAATTTT 1871
    |||
DB 3651 CAAGGAACCTAGCTCTGTGTGCGCAAGAAACAGACATTCCTAGTGTGTGCTTGT 3710
    |||
QY 1872 GAAGGACCCCACTCATCTATCTATGATGAAGTGTGTCAACACAGAACTTCTATTTTCATTGC 1931
    |||
DB 3711 TACACAGCTCATATTTTGTGTTGATGAAGCCAGTCACTGTGATACAGAAAGTGA 3770
    |||
QY 1932 AGAGCTATTTCTTGTGCGCATGAAGATGTGTCAACACAGAACTTCTATTTTCATTGC 1991
    |||
DB 3771 AAAGGTGTCCAAAGAGCCCTGGAACAAGCCAGAGAGCCGACCTGTGATGTGATTGC 3830
    |||
QY 1992 ACACAGATTGTCAACAGTGTGTGATGAGATGAATCATTTGCTTGGATCAGGTTAAGT 2051
    |||
DB 3831 TCACCGCTGTCCACATCCGAATGCAAGTCTTAATAGTGTGTTTCAGAAATGCAAGT 3890
    |||
QY 2052 AGCCGAACGTGTACCCACCATGTGTTGCTTGC 2084
    |||
```

Db 3891 CAAGAGCATGACGACGATCAGCAGCTGCTGSC 3923

## RESULT 14

US-08-784-649A-5  
Sequence 5, Application US/08784649A  
Patent No. 5830697

## GENERAL INFORMATION:

APPLICANT: Sikic, Branimir I  
APPLICANT: Chen, Gang  
TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO  
CYCLOSPORIN MODULATION  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 2200 Sand Hill Road  
CITY: Menlo Park  
STATE: CA  
COUNTRY: USA  
ZIP: 94025

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/784,649A  
FILING DATE:  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Sherwood, Pamela J  
REGISTRATION NUMBER: Reg.No. 5830697 36,677  
REFERENCE/DOCKET NUMBER: 06037/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-322-5070  
TELEFAX: 415-854-0875

## INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4264 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-784-649A-5

Query Match 6.1%; Score 143.8; DB 2; Length 4264;  
Best Local Similarity 52.9%; Pred. No. 1.1e-31;  
Matches 335; Conservative 0; Mismatches 292; Indels 6; Gaps 1;

Qy 1458 AGTCCTTAGTGAATATCCTTTGAAGTCCCTGCAGAAAGAGGCGCATTTGTAGGAG 1517  
Db 3291 AGTGTCTCAGGAGCTGAGCTGAGAGTGAAGAGGCGCAGCGCTGCTGTGGGAG 3350  
Qy 1518 TAGTGGGTGAGGAAAGACAAATAGTGAAGCTATTATTCGCTTCTATGAGCTCAAAA 1577  
Db 3351 CAGTGGCTGTGGGAAGAGACAGTGTCCAGCTCTCGAGCGGTTCTAGAACCCCTTGGC 3410  
Qy 1578 GGGTACATTATCTTCTGTGCAAAATATACAGATGTAGCTGAGAAAGCTTCGGAG 1637  
Db 3411 AGGGAAGGTCTGCTTGAATGCGCAAAATAAAGCAGTAATGTTCAAGTCCCGAGC 3470  
Qy 1638 GGCAGTGGAGGTGATACCTCAGAGATGCTCTCTTCCATATCTATTATTAACACT 1697  
Db 3471 ACACCTGGGCACTGTGTCCAGAGCCCATCTGTTTGACTGACACATTTGCGAACAAT 3530  
Qy 1698 CTTATATGAAAACATCAG-----TGCTTCACTGAGAGAGTATGCAAGTGGCAAAATT 1751  
Db 3531 TGCCTATGAGACAAACAGCGGGGTGTGTCAAGAGAGATTTGAGGGCGCAAAAGGA 3590  
Qy 1752 AGCTGGAAGTATGATGCAATTTCTTGAAATGCGCAATGATATGACCCCAAGTAGGGGA 1811  
Db 3591 GGCCTAACATACATGCTTCTCATGAGTCACTGCTTAATAATATATACCTAAAGTAGGAGA 3650

Qy 1812 ACAGAGACTCAAGCTTTACAGAGAGAGAAAAGCAAGATGACATTTGCAAGGCCATTTT 1871  
Db 3651 CAAGGAACATAGCTCTCTGTGGCCAGAAACAGCATTTGCCATCTGCTGTGT 3710  
Qy 1872 GAAGAGACCCCAAGTCACTATGATGAGTGAAGTCACTTCACTGTTAATGATTTACTGA 1931  
Db 3711 TAGACAGCTCATATTTTGTGTTTGGATGAAGCCACGCTCGATCGATACAGAAAGTGA 3770  
Qy 1932 AGACATATTTCTTGTCGCAATGAAGATGTGTGCAACACAGAACTCTATTTTCATTGC 1991  
Db 3771 AAAGTGTCCAGAGAGCTTGGACAAAGCCAGAGAGGCCGACCTGCAATTTGATTTGC 3830  
Qy 1992 ACACAGATTGTCAACAGTGTGTTGATGAGAAATCATTTGTTGATCAGGGTAAAGT 2051  
Db 3831 TCACCGCTGTCCACATCCAGATGCAAGTCACTTAATGTGTGTTTCAGATGCGCAGGT 3890  
Qy 2052 AGCCGAACGTGTACCCACCATGCTTTGCTTGC 2084  
Db 3891 CAAGAGCATGACGACGATCAGCAGCTGCTGSC 3923

## RESULT 15

US-08-181-471-2  
Sequence 2, Application US/08181471  
Patent No. 5641508

## GENERAL INFORMATION:

APPLICANT: Li, Lingna  
APPLICANT: Lishko, Valery K.  
TITLE OF INVENTION: METHOD FOR DELIVERING BENEFICIAL  
TITLE OF INVENTION: COMPOSITIONS TO HAIR FOLLICLES  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Thomas Fitting  
STREET: 12526 High Bluff Drive, Suite 300  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92130

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/181,471  
FILING DATE: 13-JAN-1994  
CLASSIFICATION: 424

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/041,553  
FILING DATE: 02-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas

REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: ANTO029P

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-792-3680  
TELEFAX: 619-792-8477

## INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4646 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: cDNA

FEATURE:  
HYPOTHETICAL: NO

ANTI-SENSE: NO

NAME/KEY: CDS  
LOCATION: 425..4267

US-08-181-471-2

Query Match 6.1%; Score 143.8; DB 1; Length 4646;



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: February 15, 2003, 03:21:35 ; Search time 121 Seconds  
(without alignments)  
9870.777 Million cell updates/sec

Title: AF133659  
Perfect score: 2345  
Sequence: 1 ATGGCGCTGCTCGCATGCA.....TATTAAAAATCATCATT 2345

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCRUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID                    | Description       |
|------------|-------|-------------|--------|-----------------------|-------------------|
| 1          | 358.6 | 15.3        | 5369   | 9 US-10-076-157-3     | Sequence 3, Appli |
| 2          | 282.6 | 12.1        | 1933   | 10 US-09-953-688A-4   | Sequence 4, Appli |
| 3          | 185.8 | 7.9         | 939    | 10 US-09-974-300-6511 | Sequence 6511, Ap |
| 4          | 184.6 | 7.9         | 1737   | 10 US-09-815-242-8403 | Sequence 8403, Ap |
| 5          | 182.4 | 7.8         | 1716   | 10 US-09-815-242-6522 | Sequence 6522, Ap |
| 6          | 165.4 | 7.1         | 3576   | 10 US-09-956-004-79   | Sequence 79, Appl |
| 7          | 158.2 | 6.7         | 1233   | 10 US-09-974-300-2176 | Sequence 2176, Ap |
| 8          | 157.8 | 6.7         | 6415   | 10 US-09-070-927A-275 | Sequence 275, App |
| 9          | 157.4 | 6.6         | 3601   | 10 US-09-880-107-2299 | Sequence 2299, Ap |
| 10         | 155   | 6.6         | 1791   | 10 US-09-070-927A-255 | Sequence 255, App |
| 11         | 154.6 | 6.6         | 9047   | 10 US-09-974-300-2172 | Sequence 2172, Ap |
| 12         | 154   | 6.6         | 9047   | 10 US-09-070-927A-12  | Sequence 12, Appl |
| 13         | 153.2 | 6.5         | 3861   | 9 US-09-938-842A-2237 | Sequence 2237, Ap |
| 14         | 152   | 6.5         | 3861   | 9 US-09-938-842A-263  | Sequence 263, App |
| 15         | 151.8 | 6.5         | 1863   | 9 US-09-938-842A-2520 | Sequence 2520, Ap |
| 16         | 151.8 | 6.5         | 2320   | 10 US-09-953-259-1    | Sequence 1, Appli |
| 17         | 149.4 | 6.4         | 4480   | 10 US-09-070-927A-320 | Sequence 320, App |
| 18         | 149.2 | 6.4         | 2681   | 10 US-09-749-340-8    | Sequence 8, Appli |
| 19         | 147.6 | 6.3         | 10391  | 10 US-09-070-927A-133 | Sequence 133, App |

|    |       |     |       |                        |                     |
|----|-------|-----|-------|------------------------|---------------------|
| 20 | 147.4 | 6.3 | 4317  | 9 US-10-044-671-1      | Sequence 1, Appli   |
| 21 | 147.2 | 6.3 | 1966  | 10 US-09-070-927A-433  | Sequence 433, App   |
| 22 | 146.4 | 6.2 | 5010  | 10 US-09-917-800A-483  | Sequence 483, App   |
| 23 | 146   | 6.2 | 3512  | 10 US-09-749-340-7     | Sequence 7, Appli   |
| 24 | 146   | 6.2 | 3912  | 10 US-09-917-800A-1560 | Sequence 1560, Ap   |
| 25 | 145.2 | 6.2 | 2066  | 10 US-09-873-409-9     | Sequence 9, Appli   |
| 26 | 145.2 | 6.2 | 2856  | 10 US-09-873-409-10    | Sequence 10, Appli  |
| 27 | 145.2 | 6.2 | 3177  | 10 US-09-873-409-12    | Sequence 12, Appli  |
| 28 | 145.2 | 6.2 | 3621  | 10 US-09-873-409-14    | Sequence 14, Appli  |
| 29 | 145.2 | 6.2 | 3702  | 10 US-09-873-409-13    | Sequence 13, Appli  |
| 30 | 144   | 6.1 | 1761  | 9 US-10-260-877-25     | Sequence 25, Appli  |
| 31 | 144   | 6.1 | 1764  | 10 US-09-815-242-6863  | Sequence 6863, Ap   |
| 32 | 144   | 6.1 | 4175  | 10 US-09-749-340-3     | Sequence 3, Appli   |
| 33 | 143.8 | 6.1 | 3860  | 10 US-09-866-866A-1    | Sequence 1, Appli   |
| 34 | 143.8 | 6.1 | 3860  | 10 US-09-866-866A-3    | Sequence 3, Appli   |
| 35 | 143.8 | 6.1 | 4643  | 9 US-10-072-621-2      | Sequence 2, Appli   |
| 36 | 143.8 | 6.1 | 8630  | 10 US-09-306-417-2     | Sequence 2, Appli   |
| 37 | 143.6 | 6.1 | 2698  | 10 US-09-749-340-5     | Sequence 5, Appli   |
| 38 | 142.8 | 6.1 | 12438 | 10 US-09-070-927A-173  | Sequence 173, App   |
| 39 | 142.6 | 6.1 | 1800  | 10 US-09-974-300-6492  | Sequence 6492, Ap   |
| 40 | 142.4 | 6.1 | 7296  | 10 US-09-070-927A-59   | Sequence 59, Appli  |
| 41 | 142.2 | 6.1 | 8630  | 10 US-09-306-417-1     | Sequence 1, Appli   |
| 42 | 139.4 | 5.9 | 1201  | 10 US-09-765-272-189   | Sequence 189, App   |
| 43 | 137.4 | 5.9 | 4189  | 10 US-09-866-866A-5    | Sequence 5, Appli   |
| 44 | 135.8 | 5.8 | 981   | 10 US-09-770-445-272   | Sequence 272, Appli |
| 45 | 135.4 | 5.8 | 1194  | 10 US-09-974-300-2187  | Sequence 2187, Ap   |

## ALIGNMENTS

RESULT 1  
US-10-076-157-3  
; Sequence 3, Application US/10076157  
; Publication No. US20030027309A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Suelberger, Harald  
; APPLICANT: Joefferen, Hans Wolfgang  
; APPLICANT: Doval, Jose Luis Revuelta  
; APPLICANT: Jimenez, Albero;  
; APPLICANT: Garcia, Maria Angeles Santos  
; TITLE OF INVENTION: Genes of purine biosynthesis from Ashbya Gossypii and the use th  
; TITLE OF INVENTION: 1n  
; TITLE OF INVENTION: microbial riboflavin synthesis  
; FILE REFERENCE: 48684DIY  
; CURRENT APPLICATION NUMBER: US/10/076,157  
; CURRENT FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: US 09/212,247  
; PRIOR FILING DATE: 1998-12-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: WordPerfect v. 6.1  
; SEQ ID NO 3  
; LENGTH: 5369  
; TYPE: DNA  
; ORGANISM: Ashbya gossypii  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 55..1482  
; NAME/KEY: CDS  
; LOCATION: 1767..3299  
; NAME/KEY: CDS  
; LOCATION: 3588..4703  
US-10-076-157-3

Query Match 15.3%; Score 358.6; DB 9; Length 5369;  
Best Local Similarity 54.1%; Pred. No. 9e-90;  
Matches 777; Conservative 0; Mismatches 649; Indels 9; Gaps 2;

CY 703 AACCTGATTTGGTTTCCCTGACGACACGCGACTTATCTAGGCTATGAC 762  
DB 1 AACCTGATTTGGTTTCCCTGACGACACGCGACTTATCTAGGCTATGAC 762

```

QY 763 AGAGAAACAAGGGATATCAGTTTGTCTGAGTCTTTGGATTAATCTTCCATC 822
Db 61 CTGGTTGAAAGATATCTTATGTCAGTGCATGGTTTTCATATACGAT 120
QY 823 ATGTTGAAGTATGCTTGTGAGTGTGTTTGTATTAACAAATGGGTCACAGTTTGT 882
Db 121 AATTTGAAATATCGATGGATATGTCATATGACATACAGTTGGTGTCTTCCGCT 180
QY 883 TTGTAACCTTGGAAACACTTGGTACATACAGACATTCACAGTTGCATCAGCGT 942
Db 181 GGTATTAACCTTTCGACATATGCTTCTTATCTCCATCTTACTTTCAGAACGAGGCTG 240
QY 943 AGACTAGATTTAGAAATAGAAACAAAGACATATGATGACAGTATCTGCTATA 1002
Db 241 CGCACACGGTTTAGCGGTATGCGAAACAGGCTGACAAATAGCGCGCTAGTGGCATG 300
QY 1003 GACTACGCTGTAATTAATGAACTGTGAAGTATTTTAATTAATGAAAGATATGACAG 1062
Db 301 GATTCCTTAATTAATTTGAAAGCTGTAAAGTATTTCAATTAACGAGAAATACCTTGCGGAC 360
QY 1063 AGATATGATGATTTTGAAGAAGTATGAGACTGCTTCAATGAAAAAGTACTCTACTG 1122
Db 361 AGATATCAACATCTCTTATGTAAGTACCGGATTCACGATTAAGGTCTCGCAATGCTG 420
QY 1123 GCTATGCTGAACCTTGTGTCAAAGTCTATTTTCACTGTGCTTAAACACTTAATGCTG 1182
Db 421 GCGTTTGTAAACACCGGCGAAGCACTAATTTTACACACTGCACTGCTCAATGATGT 480
QY 1183 CTGCGCAGTCAGGGAATTTGGCAGGTACCTTACTGTGGAGATCTAGTAATGCTGAT 1242
Db 481 ATGGCTGTAAATGGTGTATGAGAGGCTCTTCAAGTGGGATCTTGTGTTAATTAAT 540
QY 1243 GGACTGCTTTTCAAGCTTTCATTAACCCCTGAACCTTCTGGAACTGTATATAGAGACT 1302
Db 541 CAATCTGATTTCCAGCTCTCCGTCGACATAAATCTTCTGTGATGCTTACCTGATCTC 600
QY 1303 AGACAGCACTCATATATATGAAACACTTGTATCTACTAAGGTAGACACCAAT 1362
Db 601 AAGCAGTCTCTATATGATATGAACTTATTTAACTGCAAAAAAATCAGTCACAAT 660
QY 1363 AAAAGCAAGTATGCACTCTCCCTTCAGATCACCAACACAGCACTACCTGCTT 1422
Db 661 AAGAACTCCCAATGCCCCAGAACCTCACCAATACCAAAACCGTGTGATA--TTGCTTT 717
QY 1423 GATTAATGTCATTTGATATACATTTGAGGSCAGAAAAGTCTTATGTAATCTTTGAA 1482
Db 718 GAAATGTATGCTTGTGCTATGACCCGAGCGCGCTATTAACATGTTCGTTTACC 777
QY 1483 GTCCCTGCAGAAAGAAAGTGGCCATTTGAGAGGTAGTGGTCAAGGAAAAAGCAAT 1542
Db 778 ATCCAGCTGGAATGAACTGCAATAGAGCCCATTCGGGCTCGGCGAAGTCCACAT 837
QY 1543 GTGAGGCTATATTTGCTTCTATAGGCTCAAAAGGATGCAATTTATTTGCTGTCAA 1602
Db 838 TTGAAGCTGTATTAATTTCTATAGGCCCCAGCAAGTGTATCTTATGTCGCGACA 897
QY 1603 AATATCAAGATGTGAGCTGTGAAAGCTTGGAGGGGAGTGGAGTGTACTACAGAT 1662
Db 898 GATATCCGGAATTTAGACTTGTCTTCTTATAGGAAAGCTATGAGTGTGCGCCCAAGAT 957
QY 1663 GCTGCTCTTCCATTAATCTATTTATTAACAACCTTATATGAAACATCAGTGTCTCA 1722
Db 958 ACTCCTCTTCAATGACAACTGCGAGAAATGTTAAATTTGGGCAATATCAGTGTCTCT 1017
QY 1723 CTTGAGGAAGTATGCAAGTGGCAAAATAGCTGCACTTCAATGATCAATTTTTCGAATG 1782
Db 1018 GACGATGAAATTTCTAGGGCCATAGAAAAGCTCAACTCAGAAAGTACTCCAGAACCTA 1077
QY 1783 CCACATGATATGACCCCAAGTAGGGGAACGAGACTCAAGCTTTCAGAGAGAAAAAG 1842
Db 1078 CCAAAAGGCGCTTCAACCGTGTAGGGAGCGCGGTTTATATATACGCGAGATGAGAAA 1137
QY 1843 CAAGAATGCAATTTGCAAGAGCCATTTTGAAGAACCCCCAGTCAATCTATATGATA 1902

```

```

Db 1138 CAAAGGCTTGTATTTGCTCGTGTCTTTGAAAGACGCTCCGCTGATGTTTTCAGACAG 1197
QY 1903 GCTACTCATGTTAGATTCATTTACTGAAGAGACTATTTCTGTGGCCAT-----GAAG 1956
Db 1198 GCTACAAAGTCTCTGATATACACAGACAGAGGACCTTTCGACACCATTCACAGAAAC 1257
QY 1957 GATGTGTCAACACAGAACTTTATTTTCAATGTCACAGATTTGTCAACAGTGTGAT 2016
Db 1258 TTTTCTTCAATTCAAAGACAGCGTTTACGTTGCCATAGACTGCGCAATATGCTGAT 1317
QY 2017 GCAGATGAATCTTGTCTTGGATCAGGTTAAGTAAAGTAAAGTGTATCCCATCATGCT 2076
Db 1318 GCAGATGAATCATTTGTTCTTGAACAAAGTCTGTGTCGCGGAAAGGCGACACAGCTCG 1377
QY 2077 TTGCTGTAAACCCCATAGTATCTATTACAGAAATGTGCAATACAGAGAGGCC 2131
Db 1378 CTGTAGGCTACAAAGATCCCTATATCCGGGCTGTGTGGATATTCAGAAAAACC 1432

```

```

RESULT 2
US-09-953-688A-4
; Sequence 4, Application US/09953688A
; Patent No. US20020102649A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Reddy, Roopa
; APPLICANT: Gorgone, Gina
; APPLICANT: Corley, Neil C.
; APPLICANT: Azimzal, Yalda
; APPLICANT: Paterson, Chandra
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN TRANSPORT PROTEIN HOMOLOGS
; FILE REFERENCE: PR-0555 US
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/113,427
; PRIOR FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1933
; TYPE: DNA
; ORGANISM: HOMO SAPIEN
; FEATURE:
; OTHER INFORMATION: 2074412, ISUTNOT01
US-09-953-688A-4

```

```

Query Match 12.1%; Score 282.6; DB 10; Length 1933;
Best Local Similarity 50.5%; Pred. No. 9.1e-69;
Matches 743; Conservative 0; Mismatches 719; Indels 9; Gaps 2;

```

```

QY 666 TCTTTCATCTTCATTCACAACTCGATCTGGGTTTTCACCTGACAGACAGCGGAGCTT 745
Db 269 TCTTCTCCACCTGACAGCTTCACCTGACCTGCGACCTGGGCGCGCACAGGGAGG 328
QY 746 TATCTAAGCTATGACAGAAACAAAGGGTATCACTTTGTCTGTAGTCTTGTGAT 805
Db 329 TGTGGGATGCGGATGCGGACATCAGATGTACAGGGCTGTCACTACCTGCTGT 388
QY 806 TTAATCTTCTCCATCTCATGTTTGAAGTATGCTGTGACGTGCTTTTGTATTAACAAT 865
Db 389 TCAATGATATCCCAACGCTGCGCAGACATCATTTGGCATATCTACTTCAGCATGTCT 448
QY 866 GCGGTGCCAGTTGCTTGTGTAACCTTGAACCTTGTGATACATACACAGATTCACAG 925
Db 449 TCAAGCGCTGTGTTGGCTCATTTGTGTTCTGTGATAGCTTTTACTTCACTCCCTGACCA 508
QY 926 TTGAGTACACAGGTGAGAACTAGATTAGATTAAGAAATGAACAAAGCAGATATGATG 985
Db 509 TTGTGTCACTGAGTGAGAAACAAAGTTTGTCTGTCTATGAACACACAGAGAAAGCTTA 568

```

QY 986 CAGGTAATGCTGCTATAGACTCAGTGTGAATTAATGAACCTGTGAATTTTAATATG 1045  
Db 569 CCGGGGCAAGAGCTGAGTCTCTGCTAACTTCAGACGGTGAAGTATTAACAACGCGG 628  
QY 1046 AAAGATATGAACACAGAGATATGATGATTTTGAAGACGATAGAGCTGCTCATGGA 1105  
Db 629 AGAGTTAGAAAGTGAACGCTATCGAGAGGCCATCATAAATATCAGGGTTTGAAGTGA 688  
QY 1106 AAAGTACTTACTCTGCTATGCTGAACCTTGGTCAAGTCTATTTTCACTGCTGCTT 1165  
Db 689 AGTGAGGCTTCACTGCTTTTACTAATTCAGACCCAAACCTGCTGATTTGGCTCGGAC 748  
QY 1166 TAAACGCTATATAGTGTCTCCGCACTCAGGGAATGTGGCAGGTAACCTTACTGAG 1225  
Db 749 TCTGCGCGGCTCCCTGCTTTGGCGATCTTGTCTACTGACAGAAAGCTACAGGTGGG 808  
QY 1226 ATTAGTAATGTAATGAGTGTGCTTTTCACTTCTTCACTTCCCTGAACTTTCTGGAA 1285  
Db 809 ACTATGCTCTTTTGGCACTTACATTAATTCAGCTGTACATGCCCTCAATGTTTGGCA 868  
QY 1286 CTGTATATAGAGACTAGACAAAGCACTCATATGATATGAAACACTTGTTTACTCTCA 1345  
Db 869 CTAATACAGATATTCAGACCAACTTCATTCATGACATGGAACATGTTTGACTTGTGA 928  
QY 1346 AGGTAGACACCAATTTAAAGACAAAGTATGCACTCCCTTCAATCAACACACAGA 1405  
Db 929 AAGGAGACAGAAAGTAAAGACCTTCTGAGACAGGCGCTTGGCTT-----TCGA 982  
QY 1406 CAGTACCGTGGCTTTGATATGATGCACTTTGAATATATGAGGCGCAAGAACTCTTA 1465  
Db 983 AGGCGCTATGAGTTGAGAACGTCCTTACGCTATGCGGAGGCGGAGACTGTC 1042  
QY 1466 GTGAAATATCTTTGAATGCTCCCTGACAGAAAGAAAGTGGCAATGTAGAGTGTGGT 1525  
Db 1043 AGAAGTCTCTTTCACTGTATGCTGTGACAGACACTTGGCTGTGGCCCATCTGGGG 1102  
QY 1526 CAGGAAAGACAAATATGAGGCTATTTTCCGCTTATAGCTTCAAAAGGATGCA 1585  
Db 1103 CAGGAAAGACAAATTTTGGCGCTGTTCCCTTCTAAGATCAGCTCTGGCTGCA 1162  
QY 1586 TTTATCTTGTGCTGCTTCAAAATATACAGATGTGAGCTGTGAAAGCTTGGAGGCACTG 1645  
Db 1163 TCCGAATAGATGGGACAGACATTTTCAAGGAGCCAGGCTCTCTCGGCTGCAATG 1222  
QY 1646 GAGTGTACTCTCAGAGATGCTGCTCTTCAATTAATCTATTTATTAACAACCTTTATATG 1705  
Db 1223 GAGTGTGCCCCAAGACACTGCTCTTTAATGACACATGCCGACAAATATCCGTTACG 1282  
QY 1706 GAAACATCAGTCTTCACTGAGGAAGTATGCACTGGCAAAATTAGCTGCACTCATG 1765  
Db 1283 GCCGTGTCAAGCTGGGAATGATGAGGTGAGGCTGCTGCTCAGGCTGCAAGGCATTCATG 1342  
QY 1766 ATGGAATTTTCCGAATGCCATGTGATATGACACCCAAATGAGGAAACGAGATCTCAAG 1825  
Db 1343 ATGCAATTAATGCTTCTTCCGAAAGGTACAGAGACAGGTGGGAGACGGGACTGAGC 1402  
QY 1826 TTTCAGAGAGAGAAACAAAGATAGCAATTTGCAAGCCATTTTGAAGACCCCCAG 1885  
Db 1403 TGAAGCGCGGGAGAAACAGGCGCTCCGCACTTGGCCGACCATCTCAAGGCTCGGGA 1462  
QY 1886 TCATATCTATATGAAAGTACTTCACTGTTAGATTGATTAAGAGAGACTATCTTGG 1945  
Db 1463 TCATTTCTGTGATGAGGCAAGTCAAGGCTGATATCATTAATGAGGGCCATTCAGG 1522  
QY 1946 GTGCAATGAAGATGTGTCAAAACAGAACTTATTTTCACTGCAACAGATTGTCAA 2005  
Db 1523 CTCTTCCGCAAGGTGTGTCACACGCAACATCTGATGTCACACAGGCTCTCAA 1582  
QY 2006 CAGTGTGATGATGATGAAATCTTGTGATTCAGAGTAAAGTGTGCGCAAGTGTGTA 2065  
Db 1583 CTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1642  
QY 2066 CCAACATGTTTGTCTTGTCTAACCTCATATGATATTAATGAAATGTGCAATACAGA 2125

Db 1643 GACACGAGGCTGCTGTTCTC---CCGAGGTGGGTATGCTGACATGTGGCAGCTGACG 1699  
QY 2126 GCAGCGGTGTGCAGAACCATGATTAACCCAA 2156  
Db 1700 AGGACAGAGAAACCTCTGAAGACACTTA 1730

RESULT 3  
US-09-974-300-6511  
; Sequence 6511, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groch  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085.500-US  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6511  
; LENGTH: 939  
; TYPE: DNA  
; ORGANISM: Bacillus clausii  
US-09-974-300-6511

Query Match 7.9%; Score 185.8; DB 10; Length 939;  
Best Local Similarity 56.4%; Pred. No 7.4e-42;  
Matches 346; Conservative 0; Mismatches 267; Indels 0; Gaps 0;

QY 1477 TTGAAGTCCCTGCGAGAAAGAAAGTGGCCATTTGAGAGTATGAGTGTGAGGAAAGC 1536  
Db 285 TTAAACATTCACAGAGGCGAGACGGTTGCCCTTGTGGCATGAGCGGTGAGGGAAGC 344  
QY 1537 ACAATAGTGGCTATTTATTTCCCTTATGAGCTTCAAAAGGATGACATTTATCTTGTCT 1596  
Db 345 ACCCTGCTGAGCTGATACCGAGGTTTTCAGATGTTACAAAGCCGATTCGATTTGAC 404  
QY 1597 GGTCAAAATATACAAAGTGTGAGCTGGAAGCCTTGGAGGCGAGTGGAGTGTACT 1656  
Db 405 GCGAAAGACATTTGTGATCTCTGTGTGAGAGTTTGGCGCAATATTTGCAATGTTCTG 464  
QY 1657 CAGATGCTGTCTCTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1716  
Db 465 CAAATATATATCTGTTTACGACAGGTGAATGCAATTAATTAATTAATTAATTAATTA 524  
QY 1717 GCTTACCTGAGAGATGATGAGTGGCAAAATTAGCTGCACTTATGATGCAATTTCTT 1776  
Db 525 GCAACAGATACAAAGTCAATTCAGGACGCAAAAGCCAAACCCCATGATTTATTTGC 584  
QY 1777 CGAATGCAATGATATGACACCAAGTATGAGGGAACGAGGACCTTTCAGAGGA 1836  
Db 585 AACCTGAAATGCTATGATACGAAATGAGGAGGAGGAGCAAACTATCTGAGGG 644  
QY 1837 GAAAGCAAGATAGCAATTTGCAAGCCATTTTGAAGAGCCCCAGTCACTACTAT 1896  
Db 645 CAAGAGAGGAGCATTTGCAATTTGCCGATATTTTAAAGAAATCACTGTGCTATTTT 704  
QY 1897 GATGAAGTACTTCAATGCTTATGATTTGATTAAGAGACTATTTCTGCTGCAATGAAG 1956  
Db 705 GAGGAAGCAACGTCAGCGCTGATTTTGAAGCAACATTTTATCAAGAAAGCACTAGAA 764  
QY 1957 GAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2016  
Db 765 AACCTGCGCGAAGCGAGAACGTTTATGCTGACATGCTTTTGTACAACTATTAAGAT 824  
QY 2017 GCAGATGAATCATTTGCTTGTGATCAGGGTAAGTATGAGCCGAGCTGTAACCATGAT 2076

Db 825 GCCGATACGATGTGCTTATAGACGCGAAAGTAGTGAACAAGCCTCATAAAGA 884  
Qy 2077 TTGCTTGCTAAC 2089  
Db 885 TTAATGGAAGC 897

## RESULT 4

US-09-815-242-8403

; Sequence 8403, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zykkind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8403  
; LENGTH: 1737  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1737)  
US-09-815-242-8403

Query Match 7.9%; Score 184.6; DB 10; Length 1737;

Best Local Similarity 56.8%; Pred. No. 2.4e-41; Matches 340; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

Qy 1491 AGAAGAAGTGGCCATTGTAGAGTGTGGTCCAGGAAAGCACATATAGAGCT 1550  
Db 1098 AGGAAACAGTGTCTTGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1157  
Qy 1551 ATTATTTGCTTCTATGAGCTCAAAAGGTAGCAATTATCTTGTGTCAAAATATCA 1610  
Db 1158 AATACCGAGATTTAGATGTAATCTTGGCAAAATTTAATAGATGTCACAAATTA 1217  
Qy 1611 AGATGTGAGCTGGAAAGCTTCGAGGAGGAGTGGAGTGAAGTGAAGTGTCTCT 1670  
Db 1218 AGATTTTAAAGGGAAGTTTAAAGAAATCAATAGATTTGGTGAACAGGATTAATTT 1277  
Qy 1671 CTTCATATATATTTATTAACAACCTTTATATGGAACAATCAGTGTCTTCACTGAGA 1730  
Db 1278 ATTCTCCGATAGTTAAGAAATATTTTACTTGTGTCTCAACGACAAAGATGAAAG 1337  
Qy 1731 AGTGTATCAGTGGCAAAATTTAGCTGAGCTTCATGATGCAATTTCTGCAATGCCCATNG 1790

Db 1338 AGTAGTGAAGCGCGAAATAGCTTAATGACATGACTTTATTTAATGAACTTGCCACAGG 1397  
Qy 1791 ATATGACACCCAGTAGGAGGAAGAGACCTCAGCTTTACAGAGAGAAAGAAAGT 1850  
Db 1398 ATATGACACTGAGTAGGTGAACAGAGTGTAAATTTACAGTGTGTCAAAAAGATT 1457  
Qy 1851 AGCAATTGCAAGAGCAATTTTGAAGACCCCGAGTCATATCTATGATGAAGTACTTC 1910  
Db 1458 ATCGATTGCTAGAAATATTTTAAATATCCGCCAATCTTATCTCGATGAAGCAACAG 1517  
Qy 1911 ATCGTTAGATTGATCTAGTAAGAGACTATTTCTGTGGCCATGAAGATGTGTCAACA 1970  
Db 1518 TGCACTGATTTTGAAGTGAATCCATTATTCAGAAAGCATTAATGTTGAGTAAAGA 1577  
Qy 1971 CAGAACTTCATTTTCAATTTGACACAGATTGTCAACAGTGTGATGACATGAATCAT 2030  
Db 1578 TCGAACGACACTTATCGTAGGCAATCGTTGTCTATATTTACATGTGTGACAAATGT 1637  
Qy 2031 TGTCTTGATCAGGATTAAGTACCGAACTGTGTAACCATGATGTTGCTTGAAC 2089  
Db 1638 CGTAATTGAATAATGGGCAATTTGTTGAAACAGTAGCATCGTAATGATTGCAAAAC 1696

## RESULT 5

US-09-815-242-6522

; Sequence 6522, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zykkind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6522  
; LENGTH: 1716  
; TYPE: DNA  
; ORGANISM: Enterococcus faecalis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1716)  
US-09-815-242-6522

Query Match 7.8%; Score 182.4; DB 10; Length 1716;

Best Local Similarity 53.0%; Pred. No. 9.8e-41; Matches 390; Conservative 0; Mismatches 346; Indels 0; Gaps 0;

Qy 1349 TAGACACCAAAATTAAGACAAAGTATGATGATCTCCCTTCAGATCAACACAGACAG 1408  
Db 932 TGAACAAACACCGTCATCAAAAGATTGCGAGAGCTGTGTCTTCAATTTAGAAAG 991



```

; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2176
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2176

```

```

Query Match      6.7%  Score 158.2; DB 10; Length 1233;
Best Local Similarity 53.9%; Pred. No. 4.8e-34;
Matches 325; Conservative 0; Mismatches 278; Indels 0; Gaps 0;

```

```

QY 1472 TATCTTTGAAGTCCCTGAGAGAAAGAAAGTGGCCATTGTAGAGATAGTGGTCAAGGA 1531
DB 576 TTTCTTTGAAGCCGACGCGGCGAAAGCGTGGCGTGGTGGATGAGCGGAGCGGCA 635
QY 1532 AAGACAAATAGTAGGCTATTATTTCGCTTTATGAGCCTTAAAGGTTAGCATTTATC 1591
DB 636 AATCAGCGTGTGACGCTTCCGAGATTCACGATGCTCAGCCGGAAGCCTGGAGA 695
QY 1592 TTGCTGTCAAAATATACAGATGTGAGCTGGAAAGCCTTGGAGGGGAGTGGAGTGG 1651
DB 666 TCGACGAGATGATGCGGAGATATAGAGCGCGAAGCCTTAGAAACCAATCGGCATGG 755
QY 1652 TACCTCAGAGATGCTGTCTCTTCATTAATTAATTAATTAATTAATTAATTAATTA 1711
DB 756 TTTCTCAGAGATGCTTTTATTATTCAGGATCGGTAGCGAAACATATCGAGGAATC 815
QY 1712 TCAGTGTTCACCTGAGAGAGTGTATGAGTGGCAAAATTAAGTGGATTCATATGCA 1771
DB 816 CCGAAGCTTCAGTGAAGACATATTGAAGCGGCAAAAGCTGCAATGCCATGATTTTA 875
QY 1772 TTCTTCGATGCGCATGATATGACACCCAGTAGGGGAGAGAGACTCAAGCTTTTCAG 1831
DB 876 TTATGAGACTGCGGAGAGGCTATGATACAAAGSTCGAGAGAAAGGAGATCAAGCTGTCC 935
QY 1832 GAGGAGAAAGCAAGAGTAGCAATGTGCAAGAGCAATTTGAAGAGCCCGGAGCATAC 1891
DB 936 GAGGCAAAAGCAGCGGATATCAATCGCCAGGGGTCTCTGAAGAACCGCGCTTTAA 995
QY 1892 TCTATGATGAAGTACTTCTATGTTAGATTGATTAAGTAAAGACTATTTCTGTGCA 1951
DB 996 TTTTGAATGAAGCAGTCAAGGCTTGACCTTGAAGCGAGCATATATCCAGGAAGGA 1055
QY 1952 TGAAGGATGTGTCAAACACAGACTTCTATTTTCAATGSCACAGATTTGTCAACAGTG 2011
DB 1056 TGGAAAGCTCCGAAAGACCGGAGCACTTATGTTGCTCAAGACTGTGACGATTA 1115
QY 2012 TTGATGCAAGTAAATCATTTGTTGATCAGGTAAAGGTAGCCGAGCTGTACCCACC 2071
DB 1116 CCCAGCTGACAAATTCGTTGTCATGAAAGACGAGCTGTGCGGAAATCGGCAACATG 1175
QY 2072 ATG 2074
DB 1176 ACG 1178

```

```

RESULT 8
US-09-070-927A-275/c
; Sequence 275; Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunach
; Patrick J. Dillon
; Seven Barish
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA

```

```

; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 275:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6415 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 275:
US-09-070-927A-275

```

```

Query Match      6.7%  Score 157.8; DB 10; Length 6415;
Best Local Similarity 53.5%; Pred. No. 1.8e-33;
Matches 327; Conservative 1; Mismatches 283; Indels 0; Gaps 0;

```

```

QY 1477 TTTGAAGTCCCTGAGAGAAAGAAAGTGGCCATTGTAGAGTATGAGTCAAGGAAAGC 1536
DB 6041 TTTTCGTTGCAAGAGAGAAACCTTGGGATTTTGGCAAAACAGGCGCTGTAATAAC 5982
QY 1537 ACAATAGTAGGCTATTAATTTGCTTCTATGAGCCTCAAAAGGTTAGCATTTCTTGGT 1596
DB 5981 ACTATTTGAATTTGATGCTGTAATATGACCAATATCAAGGAGCATTTCTTTGGA 5922
QY 1597 GGTGAATATATACAGATGTAGGCTGGAAGCCTTGGAGGGAGTGGAGTACT 1656
DB 5921 AAACATTAACATTAATAATTAACATTAATGATGATGATGATGATGATGATGATGAT 5862
QY 1657 CAGATGCTGTCTCTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1716
DB 5861 CAGATCAATTTCTCTTTCCATGACGTTACGATTAATTAATTAATTAATTAATTAAT 5802
QY 1717 GCTTACCTGAGAGTGTATGACGTGCAAAATTAAGTGAATTCATGATGCAATTTCT 1776
DB 5801 TTGAACAAGCAGAGGTGAACAAGCGGAGCATTAATTAATTAATTAATTAATTAAT 5742
QY 1777 CGAATGCCAGATGATATGACCAAGTAGGGGAGAGAGAGCTCAAGCTTTCAGAGGA 1836
DB 5741 GCATTCCTGAGAGCTATGACAAATGTTGGGAGAGTGGCGTTTGCCTTTTCAGAGGA 5662
QY 1837 GAAAGCAAAAGTAGAATTTGCAAGAGCCATTTTGAAGAGCCCGGAGTCACTAT 1896
DB 5681 CMAAAACAAGCATTTTCTATTGACGCGGCTTGATTTGTAACAGAAATGTTAATTTTA 5622
QY 1897 GATGAAGTACTTATGATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1956
DB 5621 GACGATGCTTATCTGAGTGTGATGAGTGAAGACTGAAGAGGCGATTTATCAATCTAAA 5562
QY 1957 GATGTGCTAAACACAGAACTTATTTTCAATTCACACAGATTTGTCAACGTGGTTGAT 2016
DB 5561 GAAACGCGCCAAAGAAAACCACTATTATTACAGCGATGTTTAAGTATGATCATCAT 5502

```

| Qy | 2017 | GCAGTGAATCATGTCTTGATCAGGGTAAGCCGAACGCGTATCCACCATGTT   | 2076 |
|----|------|---|------|
| Db | 5501 | GCCTAAAGAACTTAGCTTGATGAAGGAAATATTATTAACGGGGGACGATCCAG | 5442 |
| Qy | 2077 | TTGCTTGCTAA   | 2087 |
| Db | 5441 | TTACTGGCTCA   | 5431 |

## RESULT 9

```

US-09-880-107-2299
; Sequence 2299, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ. ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2299
; LENGTH: 3924
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M23234
; US-09-880-107-2299

```

|                       |              |                   |                |             |
|-----------------------|--------------|-------------------|----------------|-------------|
| Query Match           | 6.7%         | Score 157.4       | DB 10          | Length 3924 |
| Best Local Similarity | 53.5%        | Pred. No. 1.7e+33 |                |             |
| Matches 329           | Conservative | 0                 | Mismatches 286 | Indels 0    |
|                       |              |                   |                | Gaps 0      |

|    |      |   |      |
|----|------|---|------|
| Qy | 1456 | AAAGTCCTTAGTGAATATCTTTGAAGTCCGCGAGGAAAGAAAGTGGCACTGTGAGA        | 1515 |
| Db | 1260 | AAAGTCTTGAAGGGCCTCAACCTGGAAGGTGCAGAGTGGGACACGGTGGCCCTGTGTGA     | 1319 |
| Qy | 1516 | GGTAGTGGGTCAAGGAAAAGCAACATAGTAGGCTATATTTGCTTATAGGCTCAA          | 1575 |
| Db | 1320 | AGTAGTGGCTGTGGGAAAGAGCAACGGTCCAGCTGATACAGAGGCTCTATGACCCTGAT     | 1379 |
| Qy | 1576 | AAGGTAGCATTTATCTGTGCTCAAAATATACAGATGTAGGCTGGAAAGCTTCGG          | 1635 |
| Db | 1380 | GAGGGCAAAATTAACATTGATGGGCGAGATATAGGAACCTTTAATGTAACTATCTGAGG     | 1439 |
| Qy | 1636 | AGGGCAGTGGAGTGTGTAACCTCAGATGCTGTCTCTTCATATAATCAATTTATTAAC       | 1695 |
| Db | 1440 | GAATCATTTGTGTGGTAGTACAGAGCGCGGTGTGTTTCCACCAATGTCTGAATA          | 1499 |
| Qy | 1696 | CTCTTAATAGGAAACATCACTGCTTCACTGAGAAAGTATAGCAGTGGCAAAATTAGCT      | 1755 |
| Db | 1500 | ATTTGTTATGGCCGTGGAAATGTAAACATGATGTAGATATAAAMAAGCTGTCAAGAAGCC    | 1559 |
| Qy | 1756 | GGACTTCATGATCAATTCCTTCGAATGCCACATGTATGATACCCAGTAGTAGGGAAAG      | 1815 |
| Db | 1560 | AAGCGCATGTAGTTTATCATGAATTAACCAAGAAATTTGACACCCCTGTGTGAGAGAGA     | 1619 |
| Qy | 1816 | GGACTCAAGCTTTCAGAGAGAGAAAGCAAGGTAGCAATTGCAAGGCCAATTTGAAG        | 1875 |
| Db | 1620 | GGGGCCCGAGTGTGATGGGCGAAGACAGAGATTCGCATTTGACCTGCGCCCTGTTGCG      | 1679 |
| Qy | 1876 | GACCCCCCACTACTACTATATGATGAAGTACTTCATGTTAGATTGATTACTAGGAAG       | 1935 |
| Db | 1680 | AAACCCCAAGATCTCTTCTGCTGATGTAGGCCACATGCACTTGTGACACAGAAATGTAAAGCT | 1739 |
| Qy | 1936 | ACTATTCGTGTGCCATGAGAGATGTGTCAACAACAGAACTTCTATTTTCATTTGACAC      | 1995 |

|    |      |  |      |
|----|------|--|------|
| Db | 1740 | GAGGTACAGGACGCTTGGATAAAGCCAGAGAAGGCCGAGCCACCATTTGTGATAGCAC | 1799 |
| Qy | 1996 | AGATTGTCAACAGTGGTTGATGCAGATGAATCATTTGTTGGATCAGGGTAAGGTAGCC | 2055 |
| Db | 1800 | CGACTGCTCTACGGCCGGAATGCAGATGTCATCGCTTGAGAGATGAGTATTTGTG    | 1859 |
| Qy | 2056 | GAACGTGTACCCAC   | 2070 |
| Db | 1860 | GAGCAAGAAAGCCAC  | 1874 |

## RESULT 10

US-09-070-927A-255/c  
Sequence 255, Application US/09070927A  
Patent No. US20020120116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunesh  
Patrick J. Dillon  
Steven Barash  
TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/070,927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 255:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3601 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 255:  
US-09-070-927A-255

|                       |              |                   |                |             |
|-----------------------|--------------|-------------------|----------------|-------------|
| Query Match           | 6.6%         | Score 155         | DB 10          | Length 3601 |
| Best Local Similarity | 52.8%        | Pred. No. 7.5e-33 |                |             |
| Matches 335           | Conservative | 0                 | Mismatches 300 | Indels 0    |
|                       |              |                   |                | Gaps 0      |

|    |      |   |       |
|----|------|---|-------|
| QY | 1490 | CAGGAAAGAAAGGCGCATTTGTAGAGAGTATGAGGTCAGGAGAAAGCAATATGTGAGGC   | 15449 |
| Db | 1716 | CAGAGAAACATATCCATTGTGGCGCACTGTGCGGTAAATCAACATTATGTGAATT       | 165   |
| QY | 1550 | TATTTATTCGCTTCTATGAGCCTCAAAAGGTAGCATTTATCTTGCTGGTCAAAATATAC   | 16099 |
| Db | 1656 | TGTTAAATGCGCCTTTTACGAGGTGCACCGTGGCCAAATTACGGTTATATCTCAGCATTTA | 15979 |



```

Qy 1610 AAGATGAGCCTGGAAAGCTTCGAGAGGAGTGGTACCTGAGATGCTGCC 1669
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1596 CTCAAATTAATCTGGGACAGCTTGGAAAAAGTTTGGTATGGTTCTCAAGTACTTGGC 1537
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1670 TCTTCATTAATCTATTATTAACAACCTTTATATGAGAAACATCATGCTCTCACTGAGG 1729
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1536 TATTGATAGCTCCCAATCGGGAAAAATTTAAACGTACGGAATCTTGAAAGCTCTGACGAAG 1477
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1730 AAGTATAGCAGTGGCAAAAATTAGCTGACCTTCATGATGACATTCCTGAAATGCCACATG 1789
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1476 AATATTATGAAGCCATGAAAAAGCCCATATTTTGAATTTTCCTCATGGCTTACTCTAAG 1417
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1790 GATATGACACCCCAAGTAGGGGAAACGAGACTCAAGCTTTCAGAGAGAAAAAGCAAGAG 1849
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1416 GACTGATACCGAGATGGTTGCCAAGGGGTTAAATTTCTGAGACACACACATTTAA 1357
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1850 TAGCAATTGCAAGACCATTTTGAAGAGACCCCGCAGTCACTATATGATGAAAGTACTT 1909
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1356 TGACTATTGCAACGACGATGATTAATCCGCCATGCTTATTTTATGATGAAAGCGACCA 1297
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1910 CATGTTAGATTGATTAATCTGAAAGACATCTTGTGGCCATGAAAGATGGTCAAG 1969
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1296 GTTCAGTATGATGATTAATGACGCAAAAATTCAGAGATGCTTTTACAAATGATGACTG 1237
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1970 ACAGAACTTCTATTTTCATTGACACAGATGTCAACAGTGGTGTGATGAGATGAATCA 2029
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1236 GCCGACCAAGTTTGTATTTATGCGCATCGCTTAGCAACATCAAAAGTGTGAAAAATTC 1177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2030 TTGCTTGATCAGGGTAAAGTAGCCGAAAGTGTACCCACCATGATGCTTGTCTAAC 2089
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1176 TAGTATGATGATTAATGCTCAAGTGTGAAATGCGACCCATGACGAATATTGCAAAAAG 1117
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2090 CTCATAGTATCTATTCAGAAATGTGGCATACACAG 2124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1116 AACAGGGTATTATGCTGACCTTTACGAAGCAG 1082
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-974-300-2172
; Sequence 2172, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2172
; LENGTH: 1791
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2172

Query Match 6.6%; Score 154.6; DB 10; Length 1791;
Best Local Similarity 50.9%; Pred.No.6,2e-33;
Matches 367; Conservative 0; Mismatches 354; Indels 0; Gaps 0;

```

```

Qy 1491 AGAAAAAGAAAGTGGCCATTGTAGAGTACTGGCTCAGGAAAAACAAATAGTAGGCT 1550
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1157 GGGATATCGGCTTGCGCTTGCGGACATACAGATCAGGGAAAGACGATCGCAATCT 1216
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1551 ATTAATTTGCTTCTATGAGCTCAAAAGGGTAGCATTTATCTTGCTGCTCAAAATATACA 1610
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1217 GATGACGAGGTTTATACATGATGACACAGCGGGGCGTCAAAAGTGAACGCTATACGATTAAC 1276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1611 AAGATGAGCCTGGAAAGCCCTCGAGAGGCGAGTGGAGTGTACTGAGATGCTGTCT 1670
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1277 CGATATCTCGCTTCCGATCTTGTGCGCAATACAGATGCTGTGACGACACGTTTAT 1336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1671 CTTCATTAATTAATTAATTAACAACCTTTATATGAGAAACATCATGCTTCACTGAGGA 1730
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1337 TTTTTCGGAACGATTAATGAAAAATTTGCTTTGGCCGTCGCAACCCCGCATGAGA 1396
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1731 AGTGTATGACATGGCAAAATTAAGTGAATTCATGATGCAATTTTGGATGCCACATGG 1790
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1397 AGTATGGAACACGCAAAAGCTGTGAGGCCGATTCGTTATTTTCAAGCGCTTCCGAATGG 1456
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1791 ATATGACACCCCAAGTAGGGGAAAGAGACTCAAGCTTTCAGAGAGAAAAAGCAAGAGT 1850
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1457 ATACGCCACAGAGATGGAAGAACCGGCATGATTAATCGGAGGTAGCGGACGCTCAT 1516
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1851 AGCAATTGCAAGACCATTTTGAAGAACCCCGCAGTACATCTATGATGAAGTACTTC 1910
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1517 TTCAATTTGCGAGGGCGCTTGTGCGAGATCTCGTATTTCTATCTTGATGAAGCAACGC 1576
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1911 ATCGTTAGATTGATTAATCTGAAGACTATTTCTTGTCATGAAGAGATGTGTCAACA 1970
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1577 AAGCATTGACACAGAAAACGGAAGTCAAAATCCAGCGCTTGAACACCTGCTCCATGG 1636
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1971 CAGAACTTCTATTTTCATTGACACAGATGTCAACAGTGGTGTGATGAGATGAATCAT 2030
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1637 CAGAACGGCTATTTATGATCGCTCACAGCTTTCACAGATGAAGGATCGGATTAATCT 1656
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2031 TGTCTTGATCAGGGTAAAGTAGCCGAAAGTGTACCCACCATGATGCTTGTCTAACCC 2090
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1697 GGTCTTGACACGCGCGGAAAAATPAGAAAGAAATCACAGGAATTTGATCGGGAACG 1756
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2091 T 2091
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1757 T 1757
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-070-927A-12/c
; Sequence 12, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
              Patrick J. Dillon
              Steven Barash
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655

```



|                        | 6.5%   | Score 152;   | DB 9;        | Length 3861;    |                   |
|------------------------|--------|--|--------------|-----------------|-------------------|
| Query Match Similarity | 52.5%; | Pred.  | No. 5.4e-32; |                 |                   |
| Matches                | 332;   | Conservative   | 0;           | Mismatches 300; | Indels 0; Gaps 0; |
| OY                     | 1456   | AAAGTCCTTAGTGGGAATATCCTTTGTAAGTCCCGCGAGAAGAAAGTGCCATTGTGAGA    | 1515         |                 |                   |
|                        |        |  |              |                 |                   |
| Db                     | 1150   | AAGATTCTTAAACAATTCTGCCTCTCTGTTCTCGTCGGGAAAACCAATTCCTTTGGTTGGG  | 1209         |                 |                   |
|                        |        |  |              |                 |                   |
| OY                     | 1516   | GSTAGTGGGTGACGGGAAGAAGCAAATAGTAGAGCTATTATTTCGCTTATAGAGCCTGAA   | 1575         |                 |                   |
|                        |        |  |              |                 |                   |
| Db                     | 1210   | AGCAGTGTGTTACGAAAAAACAAGTCGTTTTGCTTATAGAGCGGTTTTACGCCGGAAC     | 1269         |                 |                   |
|                        |        |  |              |                 |                   |
| OY                     | 1576   | AAGGTAGCATTTATCTTGCTGTGCAAAATATATACAAGATGTAGAGCCTGGAAAGCCTTGGG | 1635         |                 |                   |
|                        |        |  |              |                 |                   |

| Query                 | 1415  | TGGCCTTGATATATGCAATTTTGATACATTTAGAGGCCAGAAAGCTCCCTAGTGTGATAT | 1474 |
|-----------------------|-------|--|------|
| Query Match           | 6.5%  | Score 151.8; DB 9; Length 1863;                              |      |
| Best Local Similarity | 52.3% | Pred. No. 3.9e-32;   |      |
| Matches               | 336;  | Conservative 0; Mismatches 307; Indels 0; Gaps 0.            |      |

```
Db 1079 TGAATTCAAAAAGTCACTTCGCTACGAGAAAGCAAGCCGGTTATTTCCGACGTGT 1138
QY 1475 CTTTGAAGTCCCTGCAAGAAAGAAAGTGGCCATTTAGAGATAGTGGTCAAGGAAAA 1534
Db 1139 CCATTACCGCCGCGACGCGACGCGCATGCGCTTGCTGGTAATCCGCGCGGTAAAT 1198
QY 1535 GCACAAATAGTGAAGCTATTTGCTTCTATGAGCTCAAAAGGTTAGCATTTATCTTG 1594
Db 1199 CCACCTGTGTCAACCTTGTGTAGTGTGTACAAACCAACACGCGCAGCCTTGCAGTAT 1258
QY 1595 CTGTCAAAAATATCAAGATGTGAGCTGGAAGCCTTCGAGGGCAGTGGAGTGTAC 1654
Db 1259 GTGCGGTGAGATGTAAAGATCTGACTTCGAGAACTTCGGCATCCGTGGGTGTGTC 1318
QY 1655 CTCAGATGCTGTCTCTTCATATACTATTTATTAACCTCTTATATGAAAAACATCA 1714
Db 1319 TCCAGGAGCCGACGCTTGTCTCTGATTTATTCAGAAAAACATGCTTACGGTCCAG 1378
QY 1715 GTGCTTCACTGAGGAAGTGTATGCAAGTGGCAAAATTAGTGAATTGATGATCAATTC 1774
Db 1379 GTGCCACCCGCGAGAGATCATCGAAGTGGCTAAGAAAGCAAGCAATGATTCATTT 1438
QY 1775 TTCGAATGCCATGATATATACACCAAGTAGGGGAAAGAGACTCAAGCTTTCAGAG 1834
Db 1439 CCGCCTTCCCTGAAGATATGAAACCGTGTGAGAACGCGGACTCAAACTTCTGTG 1498
QY 1835 GAGAAAAGCAAAAGATGCAATTTGCAAGAGCCATTTTGAAGGACCCCGAGTCATACCT 1894
Db 1499 GCCAGAAAGACGCGCTCTCTGTGCAAGGCGCATGCTTAAGATGCCCATCTTCTGTTC 1558
QY 1895 ATGATGAAGCTACTTCACTGTTAGATTGATTAAGTGAAGACTATTTCTGTGCCATGA 1954
Db 1559 TCGATGAAGCCACTCTGCACTGATACCAAGTCTGAGCAGGCAAGTCCAAAGCGGTTGG 1618
QY 1955 AGGATGTGTCAAACACAGAACTTCTATTTTCATTGCAACAGATTGTCAAGTGTGG 2014
Db 1619 AACGCTGATGAAAAACGCACTTAATGATGCGCCACCGCTGTCCACATTCGAG 1678
QY 2015 ATGCAGATGAATCATTTGTTGATCAGGTTAGGTAGCCGA 2057
Db 1679 GCGTCATACCATGTGATCCAAAAAGGACGGGTTGAAGA 1721
```

Search completed: February 15, 2003, 05:51:32  
Job time : 148 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2003, 04:48:00 ; Search time 24 Seconds  
(without alignments)  
8105.166 Million cell updates/sec

Title: AF133659  
Perfect score: 4254  
Sequence: 1 ATGGCGCTGCTCGCATGCA.....TATTAAAAAATCATACATT 2345

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p model -DEV=xlh  
-Q=/cgr2\_1/USPTO\_spool/AF133659/runat\_10022003\_155440\_29513/app\_query.fasta\_1.2503  
-DB=SwissProt\_40 -QFMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX=blomum2 -TRANS=human40.cdi -LIST=45  
-DOCALLGN=200 -THR SCORE=pcp -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcp -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=AF133659 @CGN 1.1 25 @runat\_10022003\_155440\_29513 -NCPU=6 -ICPU=3  
-NO\_XLPLY -NO\_XMAP -LARGESQDBY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | *<br>Query Match | Length | ID | Description |
|------------|--------|------------------|--------|----|-------------|
| 1          | 3836   | 90.2             | 752    | 1  | ABC7_HUMAN  |
| 2          | 3342   | 78.6             | 694    | 1  | ABCT_MOUSE  |
| 3          | 1538.5 | 36.2             | 693    | 1  | YET1_SCHPO  |
| 4          | 1516.5 | 35.6             | 690    | 1  | ATM1_YEAST  |
| 5          | 1239.5 | 29.1             | 842    | 1  | ABCB_HUMAN  |
| 6          | 973.5  | 22.9             | 830    | 1  | HMT1_SCHPO  |
| 7          | 674.5  | 15.9             | 707    | 1  | YFIC_BACSU  |
| 8          | 673.5  | 15.8             | 707    | 1  | HLV2_ECOLI  |
| 9          | 671.5  | 15.8             | 707    | 1  | HLV2_ECOLI  |
| 10         | 664.5  | 15.6             | 715    | 1  | AB10_MOUSE  |
| 11         | 660    | 15.5             | 708    | 1  | HLVB_PASHA  |
| 12         | 654    | 15.4             | 586    | 1  | Y4GM_RHISN  |
| 13         | 651.5  | 15.3             | 707    | 1  | HLVB_ACTAC  |
| 14         | 651    | 15.3             | 582    | 1  | MSBA_ECOLI  |
| 15         | 648    | 15.1             | 707    | 1  | HLVB_PASBP  |
| 16         | 641.5  | 15.1             | 707    | 1  | RTIB_ACTPL  |
| 17         | 634    | 14.9             | 712    | 1  | CYAB_BOKEP  |
| 18         | 632    | 14.9             | 587    | 1  | MSBA_HAEIN  |

|    |       |      |      |   |             |
|----|-------|------|------|---|-------------|
| 19 | 631.5 | 14.8 | 707  | 1 | HLVB_PROVU  |
| 20 | 618.5 | 14.5 | 711  | 1 | RT3B_ACTPL  |
| 21 | 618   | 14.5 | 575  | 1 | YWJA_BACSU  |
| 22 | 613   | 14.4 | 738  | 1 | AB10_HUMAN  |
| 23 | 613   | 14.4 | 1302 | 1 | MDR4_DROME  |
| 24 | 611   | 14.4 | 614  | 1 | YAS1_HAEIN  |
| 25 | 610   | 14.3 | 631  | 1 | YC72_MYCTU  |
| 26 | 601.5 | 14.1 | 598  | 1 | Y288_THEMEA |
| 27 | 596.5 | 14.0 | 1276 | 1 | MDR2_CRIGR  |
| 28 | 592.5 | 13.9 | 717  | 1 | COMA_STRPN  |
| 29 | 586.5 | 13.8 | 762  | 1 | ABCG_MOUSE  |
| 30 | 586   | 13.8 | 762  | 1 | ABCG_RAT    |
| 31 | 585   | 13.8 | 1276 | 1 | MDR1_MOUSE  |
| 32 | 582.5 | 13.7 | 766  | 1 | ABCG_HUMAN  |
| 33 | 582.5 | 13.7 | 1278 | 1 | MDR2_RAT    |
| 34 | 579   | 13.6 | 1302 | 1 | MDR5_RAT    |
| 35 | 577   | 13.6 | 1302 | 1 | MDR5_DROME  |
| 36 | 572.5 | 13.5 | 1276 | 1 | MDR2_MOUSE  |
| 37 | 572.5 | 13.5 | 1280 | 1 | MDR1_HUMAN  |
| 38 | 572.5 | 13.5 | 1281 | 1 | MDR3_CRIGR  |
| 39 | 571.5 | 13.4 | 735  | 1 | ABCB_HUMAN  |
| 40 | 571   | 13.4 | 715  | 1 | LCNC_LACLA  |
| 41 | 570.5 | 13.4 | 589  | 1 | Y015_MYCGE  |
| 42 | 570.5 | 13.4 | 722  | 1 | MESD_LEUME  |
| 43 | 569.5 | 13.4 | 607  | 1 | HEPA_ANASP  |
| 44 | 568.5 | 13.4 | 859  | 1 | YD48_MYCTU  |
| 45 | 568.5 | 13.4 | 1279 | 1 | MDR3_HUMAN  |

ALIGNMENTS

RESULT 1  
ID ABC7\_HUMAN STANDARD: PRT: 752 AA.  
AC 075027; 075345; Q9UND1; Q9P0P1; Q9BRE1;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE ATP-binding cassette, sub-family B, member 7, mitochondrial precursor  
GN (ATP-binding cassette transporter 7) (ABC transporter 7 protein).  
GN ABCB7 OR ABC7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=98284536; PubMed=9621516;  
RA Shimada Y., Okuno S., Kawai A., Shinomiya H., Saito O., Suzuki M.,  
RA Omori Y., Nishino N., Kanemoto N., Fujiwara T., Horie M.,  
RA Takahashi E.,  
RT "Cloning and chromosomal mapping of a novel ABC transporter gene  
RT (hABC7), a candidate for X-linked sideroblastic anemia with  
RT spinocerebellar ataxia.";  
RL J. Hum. Genet. 43:115-122(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=98318631; PubMed=9653160;  
RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,  
RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,  
RA Wang Y.-X., Chen S.-J., Chen Z.,  
RT "Identification of genes expressed in human CD34(+) hematopoietic  
RT stem/progenitor cells by expressed sequence tags and efficient full-  
RT length cDNA cloning.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).  
RN [3]  
RP SEQUENCE FROM N.A., AND VARIANT ASAT MET-400.  
RX MEDLINE=99214014; PubMed=10196363;  
RA Allikmets R., Raskind W.H., Hutchinson A., Schneck N.D., Dean M.,  
RA Koeller D.M.,  
RT "Mutation of a putative mitochondrial iron transporter gene (ABCT7) in

```

RT X-linked sideroblastic anemia and ataxia (XLSA/A).";
RL Hum. Mol. Genet. 8:743-749(1999).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANT ASAT LYS-433.
RX MEDLINE=20504074; PubMed=11050011;
RA Bekri S., Kispal G., Lange H., Fitzsimons E., Tolmie J., Lill R.,
RT Bishop D.F.;
RT "Human ABC7 transporter: gene structure and mutation causing X-linked
RT sideroblastic anemia with ataxia with disruption of cytosolic
RT iron-sulfur protein maturation.";
RL Blood 96:3256-3264(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 5-752 FROM N.A.
RX MEDLINE=99098366; PubMed=9883897;
RA Csere P., Lill R., Kispal G.;
RT "Identification of a human mitochondrial ABC transporter, the
RT functional orthologue of yeast Atmtp.",
RL FEBS Lett. 441:266-270(1998).
CC -! FUNCTION: COULD BE INVOLVED IN THE TRANSPORT OF HEME FROM THE
CC MITOCHONDRIA TO THE CYTOSOL. PLAYS A CENTRAL ROLE IN THE
CC MATURATION OF CYTOSOLIC IRON-SULFUR (FE/S) CLUSTER-CONTAINING
CC PROTEINS.
CC -! SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL). *
CC -! SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (Potential).
CC -! DISEASE: DEFECTS IN ABCB7 ARE THE CAUSE OF X-LINKED SIDEROBLASTIC
CC ANEMIA WITH ATAXIA (ASAT). ASAT IS A RECESSIVE DISORDER
CC CHARACTERIZED BY AN INFANTILE TO EARLY CHILDHOOD ONSET OF
CC NONPROGRESSIVE CEREBELLAR ATAXIA AND MILD ANEMIA WITH HYPOCHROMIA
CC AND MICROCYTOSIS.
CC -! SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AB005289; BAA28861.1; -.
DR EMBL, AF038950; AAC39850.1; -.
DR EMBL, AF133659; AAD33045.1; -.
DR EMBL, AF241887; AAK20173.1; -.
DR EMBL, AF241872; AAK20173.1; JOINED.
DR EMBL, AF241873; AAK20173.1; JOINED.
DR EMBL, AF241874; AAK20173.1; JOINED.
DR EMBL, AF241875; AAK20173.1; JOINED.
DR EMBL, AF241876; AAK20173.1; JOINED.
DR EMBL, AF241877; AAK20173.1; JOINED.
DR EMBL, AF241878; AAK20173.1; JOINED.
DR EMBL, AF241879; AAK20173.1; JOINED.
DR EMBL, AF241880; AAK20173.1; JOINED.
DR EMBL, AF241881; AAK20173.1; JOINED.
DR EMBL, AF241882; AAK20173.1; JOINED.
DR EMBL, AF241883; AAK20173.1; JOINED.
DR EMBL, AF241884; AAK20173.1; JOINED.
DR EMBL, AF241885; AAK20173.1; JOINED.
DR EMBL, AF241886; AAK20173.1; JOINED.
DR EMBL, BC006323; AAH06323.1; -.
DR EMBL, AF078777; AAD47141.1; -.
DR Genew; HGNC:48; ABCB7.
DR MIM, 300135; -.
DR InterPro, IPR003593; AAA_ATPase.
DR InterPro, IPR003439; ABC_transporter.
DR InterPro, IPR001140; AbcTransportTM.
DR Pfam, PF00005; ABC_tran; 2.

```

|    |   |         |                                   |
|----|---|---------|-----------------------------------|
| DR | Pfam: PF00664; ABC membrane; 2.                                       |         |                                   |
| DR | ProDom: PD000006; ABC transport; 1.                                   |         |                                   |
| DR | SMART: SM00382; AAA; 1.   |         |                                   |
| DR | PROSITE: PS00211; ABC_TRANSPORTER; 1.                                 |         |                                   |
| KW | ATP-binding; Transmembrane; Transport; Mitochondrion; Inner membrane; |         |                                   |
| KW | Transit peptide; Disease mutation.                                    |         |                                   |
| FT | TRANSIT   | 1       | ?                                 |
| FT | CHAIN   | ?       | 752                               |
| FT |   |         |                                   |
| FT | TRANSMEM  | 260     | 280                               |
| FT | TRANSMEM  | 291     | 311                               |
| FT | TRANSMEM  | 383     | 403                               |
| FT | NP_BIND   | 505     | 512                               |
| FT | VARIANT   | 400     | 400                               |
| FT |   |         |                                   |
| FT | VARIANT   | 433     | 433                               |
| FT |   |         |                                   |
| FT | CONFLICT  | 56      | 56                                |
| FT | CONFLICT  | 141     | 141                               |
| FT | CONFLICT  | 258     | 258                               |
| FT | CONFLICT  | 271     | 276                               |
| FT | CONFLICT  | 281     | 281                               |
| FT | CONFLICT  | 290     | 290                               |
| FT | CONFLICT  | 293     | 297                               |
| FT | CONFLICT  | 315     | 315                               |
| FT | CONFLICT  | 320     | 324                               |
| FT | CONFLICT  | 346     | 346                               |
| FT | CONFLICT  | 542     | 542                               |
| QO | SEQUENCE  | 752 AA; | 82641 MW; B1FEA57ABD24FB90 CRC64; |

```

AF133659 (1-2345) x ABC7_HUMAN (1-752)
Alignment Scores:
Pred. No.:      7,96e-355      Length:      752
Score:          3836.00        Matches:      752
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    90.17%        Indels:      0
DB:             1            Gaps:        0

```

|    |     |  |     |
|----|-----|--|-----|
| Oy | 1   | ATGGCGCTGCTCGCGATGCATCTTTGGCGCTGGCGCGCGCGCTGCTTTGAAAG              | 60  |
| Db | 1   | MetAlaLeuLeuAlaMetHisSerTyrArgTyrAlaAlaAlaAlaAlaAlaPheGluLys       | 20  |
| Oy | 61  | CGCCGGCAGCTCGCGGATCTCATCGCGCTTTAGTCTCTGTAAACGGCTCAGGTCGCGAG        | 120 |
| Db | 21  | ArgArgHisSerAlaIleLeuIleIleArgProLeuValSerValSerGlySerGlyProGln    | 40  |
| Oy | 121 | TGGAGGCCACATCACTGGCGGCTTTGGGAAACCGCTCGACCTTACAGATTCACAGACTCA       | 180 |
| Db | 41  | TyrArgProHisGlnLeuGlyAlaLeuGlyThrAlaArgAlaTyrGlnIleProGlnSer       | 60  |
| Oy | 181 | TTAAAAAGTATCACATGGCAGACAGATTGGGAAAAAGCAATTCAGAGCAGTCTTAGATGCT      | 240 |
| Db | 61  | LeuLysSerIleThrTyrGlnArgLeuGlyLysGlyLysMetSerGlyGlnPheLeuAspAla    | 80  |
| Oy | 241 | GCAAGGCTCTCCAGGTATGCGCACTGATAGAAAAGAGACATGTTGGCATGCTCATGCA         | 300 |
| Db | 81  | AlaLysAlaLeuGlnValTyrProLeuIleGluLysArgThrCysTyrHisGlyHisAlaIa     | 100 |
| Oy | 301 | GGAGGAGGACCTCCACACAGACCCAAAAAGAAAGGCTTAAAGATGTGTATCTCCGGAATC       | 360 |
| Db | 101 | GlyGlyGlyLeuHisIleThrAspProLysGluGlyLeuLysAspValIleAspThrArgLysIle | 120 |
| Oy | 361 | ATAAAGCAATGCTTTCTTATGTGTGGCCCAAGACAGGCCAGATCTACAGCATGAGATT         | 420 |
| Db | 121 | IleLysAlaMetLeuSerTyrValTyrProLysAspArgProPheLeuArgAlaArgVal       | 140 |
| Oy | 421 | GCCATTTGCTGGGATTTTGGGTGGTGGCAAGGCCATGAATATGTGGTCTCCCTATG           | 480 |
| Db | 141 | AlaIleSerLeuGlyPheLeuGlyGlyAlaLysAlaMetCysIleValValProPheMet       | 160 |



QY 481 TTTAAATATGCTGATGACAGCCTCAACAGATGTCGGGAAACATGCTGAACCTGATGAT 540  
 DB 161 Pheylsrylavalalepserleuasnlglnmetserglyasmetleuasleuaserasp 180  
 QY 541 GCACCAATATGATGACACCATGGCAACAGCTTGTGATGGCTATGATGATCAAGA 600  
 DB 181 Alaprosanthrvalalathmetcalathralavaluileglytyrglyvalserhrg 200  
 QY 601 GCTGAGCTGCTTTTAAAGAGTTGAAATGCAATGCAATTTGGCAAGTGAAGCCAGAT 660  
 DB 201 AlaglyalaalaphepneasnlglnvalargasnalaValPheglylyvalalaglinsn 220  
 QY 661 TCATCCGAAATAATGCAAAATGCTTTCTCCATCTTCAACCTGAGATCTGGGTTT 720  
 DB 221 SerlleagaagllealvalysasnvalPheleuthsleuhsasnleuasplueglyphe 240  
 QY 721 CACCTGAGCAGACAGACGGAGCTTTATCTAAGCTATGACAGAGAAACAGGGGATC 780  
 DB 241 Hisleuserahglnthrglyalaleuaserlysalalleasparglythralrgglyle 260  
 QY 781 AGTTTGTCTGAGTGTCTTGGATTTAATCTTCTCCATCATGTTGAAATGATCTT 840  
 DB 261 SerphevalleuseralaleuvalPheasnleuueuproilemetPheglyvalmetleu 280  
 QY 841 GTCAGTGTGTTTGTATTAACAATGCGGTCCAGATTGCTTGGTACCCTTGAAACA 900  
 DB 281 Valserglyvalleuylrtyrlyscs6glyalaglnphealaleuvalthmleuglythr 300  
 QY 901 CTGATGATCATGACAGATTCAGAGTTGACAGTGCACGGTGGAGACATGATTTAGATA 960  
 DB 301 Leuglythrythrvalaphehthrvalalavalathralthralgthrargphargyle 320  
 QY 961 GAAATGAACAAAGAGATTAATGATGACAGTATGCTGCTATAGACTGCTGGAATTA 1020  
 DB 321 Gluwerasnlysalaspaasnaspalaglyasnalaaleaspsertleuasntyr 340  
 QY 1021 GAAATGATGAGATTTTAAATGAAGAATATGAGACAGACATGATGATTTTGG 1080  
 DB 341 Gluthrvallystyrpheasnasnlglnarglygluaglnarglyrasyrpglypheleu 360  
 QY 1081 AAGAGTATGAGACTGCTTCACTGAAAGTACTCTACTGCTGATGCTGAACTTTGGT 1140  
 DB 361 Lysthrlyrglnthralaserleuysertserthrleuualawetleuasphgily 380  
 QY 1141 CAAGTCTATTTTCAAGTGTGCTTTAAACAGCTAATGATGCTGCGCAGTCAAGGAAT 1200  
 DB 381 GluserlathlepheserValglyleuthralallemetvalleuualaserGlnglyile 400  
 QY 1201 GTGGCAGATACCTTACTGTTGAGATCTAAGTAATGATGATGAGTCTTTTCACTT 1260  
 DB 401 ValalaglythrleuthrvalglyAspleuValMetvalasnlglyleuLeuPheGlnleu 420  
 QY 1261 TCATACCCCGAATCTTTCTGGGAAGCTGATATGAGAGACTAGACAAAGCATCATGAT 1320  
 DB 421 SerleuProleuanPheleuglythValtyrarglythraaglnalaleuileasp 440  
 QY 1321 ATGAACACTTGTACTCTACTCAAGTAGACACCCAAATTAAGCAAAAGTATGCA 1380  
 DB 441 MetasnhrleuethrleuLeuValasphrglnleleuysaaplyvalMetala 460  
 QY 1381 TCTGCCCTTCAATCACAACACAGACAGCTGCGTGGCTTTGATTAATGCTATTGAA 1440  
 DB 461 SerProleuglnlethrProglnthralathrValalalpheaspaenValHIsphgln 480  
 QY 1441 TACATTGAGGGCAGAAAGCTTATGAGAAATATCCTTTGAAGTCCCTGAGAGAAAGAAA 1500  
 DB 481 TyrlleuglnlyglnlysalaleuserglylleserphelglnvalProalaglylylys 500  
 QY 1501 GTGGCCATTGAGAGGTAGTGGGTCAAGGAAACCAATAGAGAGGTATTATTTCG 1560  
 DB 501 ValalalalvalaglyserglyserglylyserthrilevalalaglyleuLeuPhearg 520  
 QY 1561 TTCTATGAGCTCAAAAGGGTAGCATTATCTTGCTGGTCAAAATATATCAAGATGTGAGC 1620

DB 521 PheylsrylProglInlysglyserllelyrleuualaglyglinsnlglnaspvalSer 540  
 QY 1621 CTGAAAGCCTTCGAGGGGAGGAGTGGATGCTACCTCAGAGATGCTGCTTCATTAAT 1680  
 DB 541 LeuglnserleuargargalavalglyvalValProglInaspalalaleuPhehsasn 560  
 QY 1681 ACTATTTATTAACAATCTTTATATGAAACATGACAGTGTCTTCACTGAGAGATGATGCA 1740  
 DB 561 ThrlelytyrasnleuLeuTyrglyAsnilleSerlaserProglIngluValTyrala 580  
 QY 1741 GTGGCAAAATTAAGTGGATCTTATGATGCAATCTTTCGAATGCCACATGATATGACACC 1800  
 DB 581 ValAlaleuValaglyleuHIsasppalalleuargmetProhIsglytyrAspThr 600  
 QY 1801 CAAGTGGGGAACGAGACTCAAGCTTTCAGAGAGAGAAAGCAAGAGATGACAAATGGCA 1860  
 DB 601 GlInvalglylnargglyleuylsleuserglyglyGlnlysglnargvalalaleala 620  
 QY 1861 AGAGCAATTTGAAGACCCCGCATATCTATGATGAGAACTTATCATGCTTATGAT 1920  
 DB 621 ArgAlalileuylsaspPropovalilleuTyraspGlnalathrsersertleuas 640  
 QY 1921 TCGATTACTGAAGACATATCTTGTGTCATGAGATGATGTCAAACAGAACTTCT 1980  
 DB 641 SerlethrGlnlythrlleleuglyalameulyaspaValVallyshsargthrser 660  
 QY 1981 ATTTTCATGCAACAGATTTGACAGGTGTTGAGAGATGAATTCATTTGCTGGAT 2040  
 DB 661 IlephellealhsargleuserthrvalValaspalaspslullellevalleuas 680  
 QY 2041 CAGGTATGAGTACCGAAGCTGTATCCACCATGTTGCTTGTCTAACCCTCATAGATC 2100  
 DB 681 GlnglylyvalalaglnarglythrhshsglyleuLeuulanspnohIsertle 700  
 QY 2101 TATTCGAATGTGGCATACACAGACAGCGGTGTGCAAGACATGATTAACCCAAATG 2160  
 DB 701 TyserGlnmettrphshsrglnserSerArgvalGlnasnhsaspsnProlystrp 720  
 QY 2161 GAAGCAAGAGAAATATATCAAGAGAGAGAAAGAAAGAAATTCACAGAAAGAAAT 2220  
 DB 721 GluAlalyssglunsnlleserlysglnuglnarglylsleuGlnuglnulle 740  
 QY 2221 GTCAATATGTGAAGGCTGTGGAACCTGTGTCG 2256  
 DB 741 ValasnserVallysglyCysglyasnCysserCys 752

RESULT 2  
 ABC7 MOUSE  
 ID ABC7 MOUSE STANDARD; PRT: 694 AA.  
 AC 061102;  
 DT 15-DEC-1998 (rel. 37, Created)  
 DT 15-DEC-1998 (rel. 37, Last sequence update)  
 DT 15-JUN-2002 (rel. 41, Last annotation update)  
 DE ATP-binding cassette, sub-family B, member 7, mitochondrial (ATP-binding cassette transporter 7) (ABC transporter 7 protein) (Fragment).  
 DE (Fragment).  
 GN ABC7 OR ABC7.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DBA/2;  
 RX MEDLINE=97288528; PubMed=9143506;  
 RA Savary S., Allikmets R., Denizot F., Luciani M.-F., Mattei M.-G., Dean M., Chimini G.;  
 "Isolation and chromosomal mapping of a novel ATP-binding cassette transporter conserved in mouse and human.";  
 RL Genomics 41:275-278(1997).  
 CC -i- FUNCTION: COULD BE INVOLVED IN THE TRANSPORT OF HEME FROM THE MITOCHONDRIA TO THE CYTOSOL. PLAYS A CENTRAL ROLE IN THE

CC MATURATION OF CYTOSOLIC IRON-SULFUR (FE/S) CLUSTER-CONTAINING  
 CC PROTEINS (By similarity).  
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U43892; AAC53152.1; -.  
 DR MGI: MGI:109533; Abcd7.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR InterPro: IPR001140; ABCtransporter.  
 DR Pfam: PF00664; ABC\_tran; 1.  
 DR Pfam: PD00006; ABC\_transport; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 KW ATP-binding; Transmembrane; Transport; Mitochondrion; Inner membrane.  
 FT NON TER 1 1  
 FT TRANSMEM 202 222 POTENTIAL.  
 FT TRANSMEM 233 253 POTENTIAL.  
 FT TRANSMEM 326 346 POTENTIAL.  
 FT NP BIND 447 454 ATP (POTENTIAL).  
 FT SEQUENCE 694 AA; 76417 MW; A7AE89EAE9AA981D CRC64;

## Alignment Scores:

Pred. No.: 4.71e-221 Length: 694  
 Score: 3342.00 Matches: 656  
 Percent Similarity: 97.41% Conservative: 20  
 Best Local Similarity: 94.52% Mismatches: 18  
 Query Match: 78.56% Indels: 0  
 DB: 1 Gaps: 0

AF133659 (1-2345) x ABC7\_MOUSE (1-694)

QY 175 GAGTCATTAAAGATGATCATGCGAGAGATTGGAAAAAGCATTCAGAGACTTTCTTA 234  
 DB 1 GlnSerLeuAArgAsnThrThrGlnGlnThrGlyLysAspAsnSerArgGlnLeuLeu 20  
 QY 225 GATGCTGAAAAGCTCTCCAGGTATGCGCCATGATAGAAAAGAGACATGTTGGCATGCT 294  
 DB 21 AspAlaThrLysAlaLeuGlnThrThrProLeuIleGlnLysArgThrCysTrpHisGly 40  
 QY 235 CATGCAAGAGAGAGACTCCACACAGACCCAAAAGAGGTTAAAGATGTTATCTGGG 354  
 DB 41 HisAlaGlyGlyGlyLeuHisThrAspProLysGlnGlyLeuLysAspValAspThrArg 60  
 QY 355 AAAATCATTAAGCATGCTTCTTATGTTGGGCCCAAGAGAGCCGATCTAGAGCT 414  
 DB 61 LysIleIleLysAlaMetLeuSerLysValTrpProGlnAspArgProAspLeuArgAla 80  
 QY 415 AGAGTTGCATTTCCTGGAGATTTTGGGTGGCAAAAGCCATTAATATGTTGTTCC 474  
 DB 81 ArgValAlaIleSerLeuGlnLysLeuGlnGlyAlaLysAlaMetAsnIleValAlaPro 100  
 QY 475 TTTCATGTTAAATATGCTGTAGACAGCTCAACAGATGTGGGAAACATGTTGAACCTG 534  
 DB 101 PheMetPheLysTrpAlaValAspSerLeuAsnGlnMetSerGlyAsnMetLeuAsnLeu 120  
 QY 535 AGTGATGACCAATATACGTGTGCAACCATGCAAGACAGCTTCTATGGCTATGCTGTA 594  
 DB 121 SerAspAlaProAsnThrValAlaThrMetAlaThrAlaValLeuIleGlyLysValAl 140  
 QY 595 TCAGAGCTGAGCTGCTTTTAAAGAAATGCAATGAGTATTTGGCAGAGTACG 654

DB 141 SerArgAlaGlyAlaAlaPhePheAsnGlnValArgAsnAlaValPheGlyLysValAla 160  
 QY 655 CAGATTCATCAATCCGAGAGATAGCCAAAATGCTTTCTCCATCTTCAACCTGATCTGTG 714  
 DB 161 GlnAsnSerIleLeuArgGlnIleAlaLysAsnValPheLeuHisLeuHisAsnLeuAspLeu 180  
 QY 715 GGTITTCACCTGACGACAGACAGCGGAGCTTTATCTAAGCTATTTGACAGAGAAACAGG 774  
 DB 181 GlyPheHisLeuSerArgGlnThrGlyAlaLeuSerLysAlaIleAspArgLysThrArg 200  
 QY 775 GGTATCAGTTTGTCCGAGGCTTTGGTATTTAATCTTCTCCATCAGTGTGAAGTG 834  
 DB 201 GlyIleSerPheValLeuSerAlaLeuValPheAsnLeuLeuProIleValPheGlnMet 220  
 QY 835 ATGCTGTGACGTGCTGTTTGTATTAACAATGCGGCTCCCGAGTTGCTTGTGAACCTT 894  
 DB 221 MetLeuValSerSerValLeuLysTrpLysCysGlnAlaGlnPheAlaLeuValThrLeu 240  
 QY 895 GGAACACTTGATGATACACAGCATTCACAGTTGCACTCACGCGTGAGAACTGATTT 954  
 DB 241 GlyThrLeuGlyAlaLysTrpThrAlaPheThrValAlaValThrArgTrpArgThrArgPhe 260  
 QY 955 AGAATAGAAATGAACAAACAGATTAATGATGACGTTAATGCTGCTTAATGACTCACTGCTG 1014  
 DB 261 ArgIleGlnMetAsnLysAlaAspAsnAspAlaGlyAsnAlaAlaIleAspSerLeuLeu 280  
 QY 1015 AATATGAACCTGAGATGTTTAAATATGAAGAATATGAGACAGACAGATATGATGA 1074  
 DB 281 AsnTrpGlnThrValLysTrpPheAsnAsnGlnLysTrpGlnAlaGlnAspGlyTrpAspGly 300  
 QY 1075 TTTTGAAGACGTATGAGACTGCTTCAATGAAAAGTACCTTCACTGCTGCTATGCTGAAC 1134  
 DB 301 PheLeuLysThrTrpGlnThrAlaSerLeuLysSerThrSerThrLeuAlaMetLeuAsn 320  
 QY 1135 TTGGTCAAACTGCTATTTTCAGTGTGCTTTAAACACTTAATATGTCGTCGCCAGTCAAG 1194  
 DB 321 PheGlyGlnAsnAlaIlePheSerValGlyLeuThrAlaIleMetValLeuAlaSerGln 340  
 QY 1195 GGAATTTGGCGAGGTACCTTACTGTTGAGATATGAAATGGGAATGGAGTACTCTTTT 1254  
 DB 341 GlyIleValAlaGlyAlaLeuThrValGlyAspLeuValMetValAsnGlyLeuLeuPhe 360  
 QY 1255 CAGCTTCATTACCCCTGAACTTTCTGGAACTGTATATAGAGACTAGACAAAGCATCTC 1314  
 DB 361 GlnLeuSerLeuProLeuAsnPheLeuGlyThrValLysArgGlnThrArgGlnAlaLeu 380  
 QY 1315 ATGATATGAACACCTTGTATTCTACTCAAGGTAGACACCCAAATTAAGACAAATG 1374  
 DB 381 IleAspMetAsnThrLeuPheThrLeuLysValAspThrArgIleLysAspLysVal 400  
 QY 1375 ATGGCATCTGCCCTTCAGATCACACACAGACCTACCGGCGCTTGATTAATGTGCAT 1434  
 DB 401 MetAlaProPheLeuGlnIleThrProGlnThrAlaThrValAlaPheAspAsnValHis 420  
 QY 1435 TTTGAATACATTGAGGCGCAAGAAAGTCTTATGTAATATCTTTGAAGTCCCTGCAGGA 1494  
 DB 421 PheGlnTrpIleGlnGlnGlnLysValLeuAsnGlnLysAlaSerPheGlnValProAlaGly 440  
 QY 1495 AAGAAAGTGGCATGTAGAGAGTAGGTGGGTCAAGAAAGACAAATAGTGAAGCTATTA 1554  
 DB 441 LysLysValAlaIleValGlyGlySerGlySerGlyLysSerThrIleValArgLeuLeu 460  
 QY 1555 TTTGGCTTCTATGAGCCTCAAAAAGGTAGCATTTATCTTGTGTCAAAATATACAGAT 1614  
 DB 461 PheArgPheTrpGlnProGlnLysGlySerLysLeuAlaGlyGlnAsnLeuGlnAsp 480  
 QY 1615 GTGAGCTGGAAGACCTTGGAGGCGAGTGGAGAGTGTACTCAGATGCTGCTCTTC 1674  
 DB 481 ValSerLeuGlnSerLeuArgArgAlaValGlyValAlaProGlnAspAlaValLeuPhe 500  
 QY 1675 CATATATCTATTATTAACAACCTCTTAATATGAGAAACATCACTGCTTCACTGAGAGAGTG 1734  
 DB 501 HisAsnThrIleLysTrpAsnLeuLeuTrpGlnLysAsnIleAsnAlaSerProGlnGlnVal 520



```

Db 166 ValLeuGlyTyRgLyPheAlaIrgLLePheSerThrValPheGlnGluLeuAraGnsSer 185
Oy 637 GATATTGGACAGTACCAGAAATTCATCCGAAAGATAGCCAAAAGTCTTTCAT 696
Db 186 ValPheAlaIleValSerIleSerIleAlaIleArgSerValSerSerAsnValIrgLnhIs 205
Oy 697 CTTCAACACCTGATCTGGGTTTTCACCTGACAGACAGACGAGGACTTATCTAAGCT 756
Db 206 LeuLeuAsnLeuAspMetCasnPheHisLeuSerIySGInThrcLySerIleThrArgAla 225
Oy 757 ATTGACAGAGAACAGAGGGTATCAGTTTGTCTGAGTCTTGAGTGGTATTAATCTTCTT 816
Db 226 MetAspArgGlyThrLyseGlyIleSerPheIleLeuSerSerMetValLeuHisIleIle 245
Oy 817 CCCATCATGTTGAAGTATGCTGTGACGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 876
Db 246 ProIleThrLeuGlnIleAlaMetValSerGlyIleLeuThrTyRgLyProSer 265
Oy 877 TTTGCTTTGTAACCTTGGAACACTTGTGATACATACACAGCATTCACAGTTGACGTACA 936
Db 266 PheSerAlaIleAlaIleAlaThrValAlaLeuTyRAlaLeuPheThrValIArgThrThr 285
Oy 937 CGGTGAGAACATGATTGATTAAGATTAAGAACAAACAGCATTAATGACGCTAATGCT 996
Db 286 SerTPArgThrValPheArgArgGlnAlaAsnAlaAlaAspSerLysAlaSerAlaAla 305
Oy 997 GGTATAGACTGCTGCTGAATTAATGAACCTGGAAGTATTAATTAATGAAGATGA 1056
Db 306 AlaIleGlySerLeuIleAsnTyRgAlaValIySThrPheAsnAsnIySerTyRgIu 325
Oy 1057 GCACAGAGATGATGATTTTGAAGAGCTGATGAGACTGCTTCAATGAAAAGTACTCT 1116
Db 326 MetSerArgTyGluLysHisLeuSerAlaTyRgIuLysAlaAsnValIySerAlaSer 345
Oy 1117 ACTGTGGCTATGCTGAATTTGGTCAAGGCTATTTTCAAGTGGCTTTAAACGCTATA 1176
Db 346 SerLeuAlaPheLeuAsnSerGlyGlnAlaIleIlePheSerThrAlaLeuThrLeuMet 365
Oy 1177 ATGGGTCCGACAGTACGAGGATTTGGGAGATACCTTACGTTGGAGATCTAATGA 1236
Db 366 MetTyRMetGlyCyAspArgGlyIleValIleThrSerAsnLeuThrValGlyAspLeuValMet 385
Oy 1237 GTGAATGACACTCTTTTCACTTTCATTAACCTTCAAGCTTCTGGAATCTATATAGA 1296
Db 386 IleAsnGlnLeuValPheGlnLeuSerIleProLeuAsnPheLeuGlySerAlaTyRArg 405
Oy 1297 GAGACTAGACAGACTCATATATGAACACCTGTTACTCTACTCAAGGTAGACACC 1356
Db 406 GluMetArgGlnAlaPheThrAspMetGluGlnLeuPheSerLeuLysArgIleAsnIle 425
Oy 1357 CAATTTAAAGACAAAGTATGGCATCTCCCTTCAAGTACACACACGACACTACCCGTG 1416
Db 426 GlnValIySGlnAlaProAspAlaArgAspLeuValLeu----LysGlyIySerIle 443
Oy 1417 GCTTTGATATATGTCATTTTGAATTAATCATTTGAGGGCCGCAAGAGCTTATAGTGAATATC 1476
Db 444 GlnPheAspAsnValHisPheSerTyRAsnProAsnArgProIleLeuAsnGlyCySer 463
Oy 1477 TTTGAAGTCCCTGACAGAAAGAAAGTGGCATTGTAAGAGTGAAGTGGCTACAGGAAAGC 1536
Db 464 PheAsnIleProAlaGlyAlaIyAlaValAlaPheValAlaIyAlaSerGlyCySerGlyIySer 483
Oy 1537 ACAATAGTAGGCTATTTATTTGCTCTTCTGAGCCCTCAAAAGGTAGCATTTATCTTCTG 1596
Db 484 ThrIleLeuArgLeuLeuAspThrPheTyRAspThrAspSerGlyLysIleLeuIleAsp 503
Oy 1597 GGTCAAAATATACAGATGATGACCTGGAAGCTTCGAGAGGAGTGGAGTGGTATCCT 1656
Db 504 AsnGlnAlaArgLeuAspGlnIleThrIleuAsnSerIleuArgLysAlaIleGlyAlaIyPro 523
Oy 1657 CAGATGCTGTCCTCTTCATATATATATTTATTAACACCTCTTATAGGAACATCACT 1716
Db 524 GlnAspThrProLeuPheAsnAspThrIleLeuTyRAsnIleGlyTyRcIyLysAsnProLys 543

```

```

Oy 1717 GCTTCACTGAGAGAGTATGACAGTGAACAAATTAAGTGGACTTCAATGATTCAT 1776
Db 544 AlaSerAsnAspGluIleValGlnAlaIleLysAlaIyAlaIySerIleHisAspIleIleGlu 563
Oy 1777 CGAATGCCACATGATATGACACCCAAAGTAGGGGAACGAGACTCAAGCTTTCAGAGGA 1836
Db 564 SerPheProGluGlyTyRgInThrcLyValGlyGluArgGlyLeuMetIleSerGlyIy 583
Oy 1837 GAAGACAAAGATGAAATTCATGACAGGACCATTTGAAAGACCCCGACATCATCTAT 1896
Db 584 GluIySGlnArgLeuAlaValSerArgLeuLeuLeuLysAsnProGluIleLeuPhe 603
Oy 1897 GATGAGCTACTCATGCTGATTCATTCATTAATGAGACATTAATGAGTATCTTGTCATCAAG 1956
Db 604 AspGlnAlaThrSerAlaLeuAspThrAsnThrcLysArgAlaLeuLeuAsnIleAsn 623
Oy 1957 GATGTCGTCAA-----CACAGAACTTCTATTTTCATTCACACAGATTCACAAGCTG 2010
Db 624 AspLeuIleLysGlySerHisIleYThrSerValPheIleAlaHisArgLeuArgThrIle 643
Oy 2011 GTTGATGACGATGAATCATTCATGCTGTGATGACGGGTAAAGTACCGCAACGTGTAACCCAC 2070
Db 644 LysAspCyAspIleIlePheValLeuGlnLysGlyArgValAlaGlnGlnGlySerHis 663
Oy 2071 CATGGTTTGCTGCTTAACCTTCATATGATTCATTCAGAAATGTGGCATACAGAGCAGC 2130
Db 664 GluGlnLeuMetAlaLys---AsnSerValTyRThrSerMetTrpHisSerGlnGlySer 682

RESULT 4
ID ATMI YEAST STANDARD; PRG; 690 AA.
AC P40416;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transporter ATMI, mitochondrial precursor.
GN ATMI OR MDY OR YMR301C OR YMR952.03C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JK9-3D ALPHA;
RX MEDLINE=95129546; PubMed=7828591;
RA Leighton J., Schatz G.;
RT "An ABC transporter in the mitochondrial inner membrane is required for normal growth of yeast."
RL EMBO J. 14:188-195(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98089018; PubMed=9428742;
RA Kispal G., Csere P., Gulard B., Lili R.;
RT "The ABC transporter Atm1p is required for mitochondrial iron homeostasis."
RL FEBS Lett. 418:346-350(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Connor R., Churcher C.M., Barrel J.B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLE TRANSPORTER FOR A YET UNKNOWN SUBSTRATE.
CC REQUIRED FOR MITOCHONDRIAL IRON HOMEOSTASIS.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial inner membrane.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib.ch](mailto:license@sib.ch)).

|    |                             |   |
|----|-----------------------------|---|
|    | CC                          | - - - - -   |
| DR | EMBL; X82612;               | CAA57938.1; -.  |
| DR | EMBL; X81715;               | CAA57359.1; -.  |
| DR | EMBL; Z49212;               | CAA89134.1; -.  |
| DR | SGD; S0004936;              | ATM1.   |
| DR | InterPro; IPR003593;        | AAA ATPase.   |
| DR | InterPro; IPR003439;        | ABC transporter.  |
| DR | InterPro; IPR01140;         | ABCTransprtTm.  |
| DR | Pfam; PF00005;              | ABC_tran; 1.  |
| DR | Pfam; Pf00664;              | ABC_membrane; 1.  |
| DR | ProDom; PD000006;           | ABC_transportr; 1.  |
| DR | SMART; SM00382;             | AAA_1.  |
| DR | PROSITE; PS00211;           | ABC_TRANSPORTER; 1.   |
| KW | AAP-binding; Transmembrane; | Transport; Mitochondrion; Inner membrane;                         |
| KW | Transit peptide.            |   |
| FT | TRANSIT                     | 1   |
| FT | CHAIN                       | ? 690   |
| FT | TRANSMEM                    | 112 132   |
| FT | TRANSMEM                    | 152 172   |
| FT | TRANSMEM                    | 229 249   |
| FT | TRANSMEM                    | 254 274   |
| FT | TRANSMEM                    | 348 368   |
| FT | TRANSMEM                    | 375 395   |
| FT | NP_BIND                     | 469 476   |
| FT | CONFIDENCE                  | 27 27   |
| QO | SEQUENCE                    | 690 AA; 77522 MW; 00975781IDJ31343 CRC64;<br>S->SRHS (IN REF. 1). |

**Alignment Scores:**

|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 3.07e-96 | Length:       | 690 |
| Score:                 | 1516.50  | Matches:      | 304 |
| Percent Similarity:    | 67.83%   | Conservative: | 122 |
| Best Local Similarity: | 48.41%   | Mismatches:   | 169 |
| Query Match:           | 35.65%   | Indels:       | 33  |
| GB:                    | 1        | Gaps:         | 7   |

AF133659 (1-2345) x ATM1 YEAST (1-690)

|    |     |   |       |
|----|-----|---|-------|
| QY | 355 | AAATCATTAAGAAAGCATGCTTCTTAATGTCGGCCAAAGACAGGCCAGATCTACAGACT       | 41.4  |
| Db | 90  | LysIleuLeuYsAspLeuPheArgTrpIleTrpProYsGlyAsnAsnLysValAlaGlyLe     | 109.0 |
| QY | 415 | AGAGTTGCCATTTGCTGGGATTTTGGTGTGTCCAAAGGCCATGAATATATGTGTTCCC        | 47.4  |
| Db | 110 | ArgValIleuIleAlaLeuGlyLeuIleuIleSerXAlaLysIleLeuAsnValGlnValPro   | 12.9  |
| QY | 475 | TTCATGTTTAAATATGCTGTATGACAGCCTCAACCAAGATGTCCGGAAACATGCTGAACCTG    | 53.4  |
| Db | 130 | PhePhePheYsGlnThrIleAspSer-----MetAsnIle                          | 14.0  |
| QY | 535 | AGTATGCACCAAAATATACAGTGTGCACACATGCGAACAC-----GCAGTTTCGATGGCTAT    | 58.8  |
| Db | 142 | AlaTrpAspAspProThrValAlaIleuProAlaAlaIleGlyLeuThrIleLeuCysTyr     | 161.0 |
| QY | 589 | GCTGTATCAAGAGCTGTGAGCTGCTTTTATTAACGAAGTTCGAAATGACAGATTTTGGCAAG    | 64.8  |
| Db | 162 | GlyValAlaArgPheGlySerValIleuPheGlyIleuLeuArgAsnAlaValAlaPheAlaLys | 181.0 |
| QY | 649 | GTACCCGAGAATTCATATCCGAGAATATGCCCCAAATATCTTTCATCCATCTTCCACACCTG    | 70.8  |
| Db | 182 | ValAlaGlnIleAsnAlaIleArgThrValSerIleuGlnThrPheGlnIleIleMetLysLeu  | 201.0 |
| QY | 709 | GATTCGGGTTTTCACCTGAGCAGACACAGACGGAGCTTATTAAGCTATTAAGCTATTCACAGAGA | 76.8  |
| Db | 202 | AspLeuIlyTrpHisLeuSerArgGlnThrGlyIleuThrArgAlaMetAspArgGly        | 221.0 |
| QY | 769 | ACAAGGGGATACAGTCTTGTCCAGAGCGTTGGTATTAATCTTCTCCACATATGTT           | 82.8  |
| Db | 222 | ThrYsGlyIleSerGlnValLeuThrAlaMetValPheHisIleIleProIleSerPhe       | 241.0 |
| QY | 829 | GAACTGATGCTTTCAGTGTGTTTATTAACAAAGCGGTGCGCCAGTTGCTTGGTGA           | 88.8  |

|    |      |  |  |     |
|----|------|--|--|-----|
| Db | 242  | GIUIIeSerValVal  | CysGlyIleLeuThrTyrGlnPheGlyAlaSerPheAlaIle | 261 |
| QY | 889  | ACCCGTGGAACTTGGTATACATACAGACGATTCACAGTTCAGACGAGGAGAACT           | 948  |     |
| Db | 262  | ThrPheSerThrMetLeuLeuTyrSerIlePheThrIleLeuThrThrAlaTrpArgThr     | 281  |     |
| QY | 949  | AGATTTAGAATAGAAATGAACAAGACAGATATGATGACGGTAACTCGCTATAGACTCA       | 1008                                       |     |
| Db | 282  | HisPheArgArgPheAlaIleAsnLysAlaAspAsnLysAlaAlaSerValAlaLeuAspSer  | 301  |     |
| QY | 1009 | CTGCTGAATTATGAAACTGTGAAAGTATTTTATATATGAAAGATATGAAACACAGATAT      | 1068                                       |     |
| Db | 302  | LeuIleAsnPheGluAlaValLysTyrPheAsnAsnGluLysTyrLeuAlaAspLysTyr     | 321  |     |
| QY | 1069 | GATGCAATTTTGAAGACGTATGAGACGCTGTCATGAAAGATACCTTATGCTGACTGATG      | 1128                                       |     |
| Db | 322  | AsnGlySerLeuMetAsnTyrArgAspSerGlnIleLysValSerGlnSerIleuAlaPhe    | 341  |     |
| QY | 1129 | CTGAACCTTGGCTCAAGTGCATATTTTCAGTGTCCGTTTAAACAGCTATATATGGTGTCCGC   | 1188                                       |     |
| Db | 342  | LeuAsnSerGlyGlnAsnLeuIlePheThrThrAlaLeuThrAlaMetCysThrMetGly     | 361  |     |
| QY | 1189 | AGTACAGGAATTGTGCGACGAGTACCTTACTGTGGAGATCTATGATATGATGATGACTG      | 1248                                       |     |
| Db | 362  | CysThrGlyValIleGlyGlyAsnLeuThrValGlyAspLeuValLeuIleAsnGlnLeu     | 381  |     |
| QY | 1249 | CTTTTTCAGCTTCATATACCCCGAACTTTCGGGAACTGTATATATAGACACTAGACAA       | 1308                                       |     |
| Db | 382  | ValPheGlnLeuSerValProlAsnAsnPheLeuGlySerValTyrArgAspLeuLysGln    | 401  |     |
| QY | 1309 | GCACCTCATAGATATGAACACCTTGTTACTCTACATCAAGATGAGACCAATTAATTAAC      | 1368                                       |     |
| Db | 402  | SerLeuIleAspMetGlnThrLeuPheLysIleuAsnGlyLysAsnGluValLysIleLysAsn | 421  |     |
| QY | 1369 | AAAGTATGGCATCTCCCTTCATGATCACCAAGACAGCT-----ACGCTGACCTTT          | 1422                                       |     |
| Db | 422  | -----AlaGluArgProlLeuMetLeu---ProGluAsnValProTyrAspIleThrPhe     | 438  |     |
| QY | 1423 | GATATATGCACTTTTGAATCATTTAGAGGGCCGAAAGTCCTATAGCAATATCTTTGAA       | 1482                                       |     |
| Db | 439  | GluAsnValThrPheGlyTyrHisProAspArgLysIleLeuLysAsnAlaSerPheThr     | 458  |     |
| QY | 1483 | GTCCCTCAGAGAAAGAAAGTGGCCATTTAGAGAGGATGAGGCTCAGGGAAAGACATTA       | 1542                                       |     |
| Db | 459  | IleProAlaGlyTrpLysThrAlaIleValIleGlySerSerGlyLysSerThrIle        | 478  |     |
| QY | 1543 | GTGAGCGTATATTTTGGCTTCTATAGAGCTCAAAAGGGTACATTTATCTTGCTGTCA        | 1602                                       |     |
| Db | 479  | LeuLysLeuValPheArgPheTyrAspProGluSerGlyArgIleLeuIleAsnGlyArg     | 498  |     |
| QY | 1603 | AATATCAAGATGTGAGCCCTGGAAGCCTTGGAGGGCAGTGGAGTGGTACTTCAGAT         | 1662                                       |     |
| Db | 499  | AspIleLysGlnTyrAspIleAspAlaLeuArgLysValIleGlyValValProGluAsp     | 518  |     |
| QY | 1663 | GCTGTCCCTTCCATATATCTATTTTCAACCTTTATATGAAACATCAGTGCCTCA           | 1722                                       |     |
| Db | 519  | ThrProLeuPheAsnAspThrIleTrpGluAsnValLysPheGlyArgIleAspAlaThr     | 538  |     |
| QY | 1723 | CTTGAGAGAGTATGACAGTGGCAAAATTTAGCTGTGACATGATGCAATTTCTTGATG        | 1782                                       |     |
| Db | 539  | AspGluGluValIleThrValValGluLysAlaGlnLeuAlaProLeuIleLysLysLeu     | 558  |     |
| QY | 1783 | CCACATGATATGACACCCCAAGTAGGGGAAACAAGATCAGCTTTCAGAGAGGAGAAAG       | 1842                                       |     |
| Db | 559  | ProGlnIlePheAspThrIleValIleGlyLysArgGlyLeuMetIleSerGlyGlyGluLys  | 578  |     |
| QY | 1843 | CAAGAGTAGCAATTTGCAAGAGCATTTTATGAGACCCCCAGTCACTATCTATAGATGA       | 1902                                       |     |
| Db | 579  | GlnArgLeuAlaIleAlaArgValLeuLeuLysAsnAlaArgIleMetPhePheAspLys     | 598  |     |
| QY | 1903 | GCTATCTTCATCGTTAATTCGATTACTGGAAGACCTATTTCTTGTCATCAAGAT--         | 1959                                       |     |





QY 223 GGACAGTCTTGAATGCTCCAAAGGCTCTCCAGGTATGCCATGTATGAAGAGACA 282  
|||  
Db 184 ValGlnPhe-----SerLeuTrpValLeu----- 191  
QY 283 TGTGGCATGTCATGCAGAGGAGACTCCACACA----- 318  
Db 192 -----ArgTyrValValSerGlyGlyLeuPheValLeuGlyLeuTrpAlaProGlyLeu 209  
QY 319 -----GACCCAAAGAGAGGTTAAAGATGTGATCTCGGAAA 357  
Db 210 ArgProGlnSerTyrThrLeuGlnValHisGlnGluAspGlnAspValGluArgSerGln 229  
QY 358 ATC-----ATRAAAGCAATG 372  
Db 230 ValArgSerAlaIaGlnGlnSerThrTrpArgAspPheGlyArgLysLeuArgLeuLeu 249  
QY 373 CTTTCTATGTCGTCGCCAAAGACAGCCAGATCTACAGCTAAGTTGGCTTTCGCTG 432  
|||  
Db 250 SerGlyTyrLeuTrpProArgGlySerProAlaLeuGlnLeuValValLeuIleCysLeu 269  
QY 433 GGATTTTGGGTGGTGCAGAGGCGCATGATATGTGGTCCCTTCATGTTAAATATGCT 492  
|||  
Db 270 GlyLeuMetGlyLeuGluArgAlaLeuAsnValLeuValProIlePheTyrArg----- 287  
QY 493 GTAGACAGCCTCAACAGATGTCGGAAACATGCTGAACCTG-----AGTATGACACA 546  
|||  
Db 288 -----AsnIleValAsnLeuLeuThrGluLysAlaPro 298  
QY 547 ---AATACAGTGCACCATGCGACACACA-----GTTTCGATTTGGCTAT 588  
|||  
Db 299 TrpAsnSerLeuAlaTrpThrValThrSerTyrValPheLeuLysPheLeuGlnGly 318  
QY 589 GGTGTATCAAGAGCTGAGCTGCTTTTATACAGATGTCGAATGACGATTTGGCAAG 648  
|||  
Db 319 GlyThrGlySerThrGly-----PheValSerAsnLeuArgThrPheLeuTrpIleArg 336  
QY 649 GTAGCCCAAGATTCATCCGAGAGATAGCCAAATATGCTTTCTCATCTTCACCAACTG 708  
|||  
Db 337 ValGlnGlnPheThrSerArgValGluLeuLeuIlePheSerHisLeuHisGlnLeu 356  
QY 709 GATTTGGGTTTTCACCTGAGACACAGACGGAGCTTATGTAAAGCTATGACAGAGA 768  
|||  
Db 357 SerLeuArgTrpHisLeuGlyArgArgThrGlyValLeuArgIleAlaAspArgGly 376  
QY 769 ACAAGGGGTATGAGTTTGTGCTGAGTGTGGTATTTATATCTTCCCATCATGCTT 828  
|||  
Db 377 ThrSerSerValThrGlyLeuLeuSerTyrLeuValPheAsnValIleProThrLeuAla 396  
QY 829 GAAGTATGCTGTGTCAGTGTGTTTGTATTCAAATGC-----GGTCCCAAGTTTGT 882  
|||  
Db 397 AspIleIleIle-----GlyIleIleTyrPheSerMetPhePheAsnAlaTrpPheGly 414  
QY 883 TTGGTAAACCTTGCAGAACCTTGTATCATACAGACATTCACAGTTGACGTACACGGTGG 942  
|||  
Db 415 LeuIleValPheLeuCysMetSerLeuTyrLeuThrIleValIleValThrGluTrp 434  
QY 943 AGAAGTATGATTTGAATGAAGAAATGACAGACAGATATGAGGTAATGCTGCTATA 1002  
|||  
Db 435 ArgThrLysPheAspArgAlaMetAsnThrGlnGluAsnAlaThrArgAlaArgAlaVal 454  
QY 1003 GACTCATGCTGATTTGAATAACTGTGAAGTATTTAATAATGAAGAAATGAGACAG 1062  
|||  
Db 455 AspSerLeuLeuAsnPheGluThrValLysSerTyrAsnAlaGlnSerTyrGluValGlu 474  
QY 1063 AGATATGATGATTTTGAAGACGTATGACAGCTGCTTCATGAAAGTACCTTACTCTG 1122  
|||  
Db 475 ArgTyrArgGluAlaIleIleLysTyrGlnGlyLeuGluTrpLysSerSerAlaSerLeu 494  
QY 1123 GCATAGCTGAACCTTGGTCAAGGTCTAATTTTCAGTGTGGTTTAAACGATATATGCTG 1182  
|||  
Db 495 ValLeuLeuAsnGlnThrGlnAsnLeuValIleGlyLeuGlyLeuLeuAlaGlySerLeu 514

QY 1183 CTGCGCCAGTCAAGGAATTTGGCGAGGTACCTTACTTGTGGAGATCTAGTAATGCTGAAT 1242  
|||  
Db 515 LeuCysAlaTyrPheValIthrGlnGlnLysLeuGlnValGlyLysPyrValLeuPheGly 534  
QY 1243 GACGTGCTTTTTCAGCTTTCATTAACCCCTGACCTTCTGGAACTGTATATAGAGACT 1302  
|||  
Db 535 ThrTyrIleLeuGlnLeuTyrMetProLeuAsnTrpPheGlyThrTyrTyrArgMetIle 554  
QY 1303 AGCAAGACACTCATGATATGAACACCTGTTTACTTACTTACAGGTAGACACCCAAAT 1362  
|||  
Db 555 GlnThrAsnHelleAspMetGluAsnMetPheAspLeuLeuLysGlnGluVal 574  
QY 1363 AAAGACAAAGTATGATGATGATCCCTTCATGATCACAACACAGCAGCTACCGTGGCTTT 1422  
|||  
Db 575 LysAspLeuProGlyAlaGlyProLeuArgPhe-----GlnLysGlyArgIleGluPhe 592  
QY 1423 GATTAATGATATTTGAATCATTCATTGAGGGCGACAAAGTCTTGTGTAATTCCTTTGAA 1482  
|||  
Db 593 GluAsnValHisPheSerTyrAlaAspGlyArgGluThrLeuGlnAspValSerPheThr 612  
QY 1483 GTCCCTGCAGAAAGAAAGGCGCATTTGTAGAGGTAGGTCAGAGGAAAGACACATA 1542  
|||  
Db 613 ValMetProGlyGlnThrLeuAlaLeuValGlyProSerGlyAlaGlyLysSerThrIle 632  
QY 1543 GTAGAGCTATTAATTTGCTTCTATGAGCTCAAAAGGATGACATTATCTTGCTGTCAA 1602  
|||  
Db 633 LeuArgLeuLeuPheArgPheTyrAspIleSerSerGlyCysIleArgIleAspGlyGln 652  
QY 1603 AATATACAAATGTGAGCTCGGAAAGCTTCGAGAGGCGAATGGAGATGATACCTCAGAT 1662  
|||  
Db 653 AspIleSerGlnValIthrGlnAlaSerLeuArgSerHisIleGlyValValProGlnAsp 672  
QY 1663 GCTGCTCTTCCATTAATCTATTTATTCACACCTCTTAATGGAACACAGTCTTCA 1722  
|||  
Db 673 ThrValLeuPheAsnAspThrIleAlaAspAsnIleArgTyrGlyArgValIthrAlaGly 692  
QY 1723 CTTGAGAACTGTATGACGTGGCAAAATTAAGCTGACCTTATGATGATCTTTCGATG 1782  
|||  
Db 693 AsnAspGluValGlnAlaAlaIaGlnAlaIaGlyIleHisAspAlaIleMetAlaPhe 712  
QY 1783 CCAATGATATATGACACCCCAAGTGGGAAACGAGACTCAAGCTTTTACAGAGAGAAAG 1842  
|||  
Db 713 ProGlnGlyTyrArgThrGlnValGlyGluArgGlyLeuLysLeuSerGlyGlyGlnLys 732  
QY 1843 CAAGAAGTACCAATTTGCAAGACCATTTTGAAGACCCCGACATCTCTATGATGAA 1902  
|||  
Db 733 GlnArgValAlaIleAlaAspThrIleLeuLysAlaProIlyIleIleLeuAspGlu 752  
QY 1903 GCTACTTCATCGTTAGATTCGATTACTGAAGACATATCTTGTCGACATGAAGATGTG 1962  
|||  
Db 753 AlaThrSerAlaLeuAspThrSerAsnGluArgAlaIleGlnAlaSerLeuAlaLysVal 772  
QY 1963 GTCAACACAGAACTTCTATTTTCATTCGACACAGATTTGCAACAGTGGTTGATGCAT 2022  
|||  
Db 773 CysAlaAsnArgThrThrIleValIaIaHisArgLeuSerThrValIaIaAsnAlaAsp 792  
QY 2023 GAATATGCTTCTGATGACGGTAAAGTACCGCAAGTGTGATCCCAACATGCTTGTGCTT 2082  
|||  
Db 793 GlnIleLeuValIleLysAspGlyCysIleValGluArgGlyLysHisGluAlaLeuLeu 812  
QY 2083 GCTAACCTCATGATCTCATTCAGAAATGTGGCATATACAGACAGCGCTGTGCAGAAC 2142  
|||  
Db 813 SerArg---GlyLysAlaTyrAlaAspMetTrpGlnLeuGlnGlnGlyGlnGluThr 831  
QY 2143 CATGATTAACCCCAA 2157  
|||  
Db 832 SerGluAspThrLys 836  
RESULT 6  
HMT1\_SCHPO  
ID HMT1\_SCHPO STANDARD: PRT: 830 AA.  
AC 002592, 013675, 090QW7, 090S13,  
01-JUL-1993 (Rel. 26, Created)



DT 01-JUN-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Heavy metal tolerance protein precursor.  
 GN HMT1 OR SPC74.08C OR SPC737.09C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 CX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SP223;  
 RX MEDLINE=93101038; PubMed=1396551;  
 RA Ortiz D.F., Kreppel L., Speiser D.M., Scheel G.,  
 RA Ow D.W.;  
 RT "Heavy metal tolerance in the fission yeast requires an ATP-binding  
 RT cassette-type vacuolar membrane transporter.";  
 RL EMBO J. 11:3491-3499(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21846401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros Y., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Hoiroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA Jaitis K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Neale S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voelckardt G., Aert R., Robben J., Grymopre B.,  
 RA Weltens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moesli D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Drenth S., Gloux S., Leleau V., Motier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Roebert M., Gallard C., Tallada V.A., Garzon A., Rhode G.,  
 RA Daga R.R., Cruzado L., Jimenez U., Sanchez M., del Rey F., Benito G.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- FUNCTION: INVOLVED IN METAL TOLERANCE. PROBABLY INVOLVED IN THE  
 CC TRANSPORT OF METAL-BOUND PHYTOCHELATINS. COMPARTMENTALIZES  
 CC CAMIUM WITHIN VACUOLES, THEREBY PROTECTING CELLS FROM CAMIUM  
 CC TOXICITY.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Vacuolar.  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL1005; CAA78419.1; -;  
 CC EMBL; AL031543; CAA20838.1; -;  
 CC EMBL; AL031546; CAA20865.1; -;  
 CC PIR; S25198; S25198.  
 CC HSP; P13569; INDB.  
 CC InterPro; IPR003593; AAA\_ATPase.  
 CC InterPro; IPR003439; ABC\_transport.  
 CC InterPro; IPR001140; ABCtransportTM.  
 CC Pfam; PF00005; ABC\_tran.1.  
 CC Pfam; PF00664; ABC\_membrane.1.

| DR | Prodom; PD000006; ABC transporter; 1.                        |         |           |                                     |
|----|--|---------|-----------|-------------------------------------|
| DR | SMART; SM00382; AAA; 1.                                      |         |           |                                     |
| DR | PROSITE; PS00211; ABC_TRANSPORTER; 1.                        |         |           |                                     |
| KW | Transport; Transmembrane; Signal; ATP-binding; Glycoprotein; |         |           |                                     |
| KW | Cadmium resistance.  |         |           |                                     |
| FT | SIGNAL   | 1       | 27        | HEAVY METAL TOLERANCE PROTEIN.      |
| FT | CHAIN  | 28      | 830       | POTENTIAL.                          |
| FT | TRANSMEM   | 10      | 31        | POTENTIAL.                          |
| FT | TRANSMEM   | 52      | 71        | POTENTIAL.                          |
| FT | TRANSMEM   | 88      | 108       | POTENTIAL.                          |
| FT | TRANSMEM   | 126     | 146       | POTENTIAL.                          |
| FT | TRANSMEM   | 403     | 423       | POTENTIAL.                          |
| FT | TRANSMEM   | 490     | 511       | POTENTIAL.                          |
| FT | NP BIND  | 617     | 624       | ATP (POTENTIAL).                    |
| FT | CAEBOHYD   | 150     | 150       | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CAEBOHYD   | 350     | 350       | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CONFLICT   | 79      | 79        | R -> A (IN REF. 1).                 |
| FT | CONFLICT   | 439     | 439       | S -> T (IN REF. 1).                 |
| FT | CONFLICT   | 812     | 812       | A -> R (IN REF. 1).                 |
| SO | SEQUENCE   | 830 AA; | 93993 MW; | 909FBD10D51F50A9 CnC64;             |

  

| Alignment Scores:                      |  |               |     |
|--|--|---------------|-----|
| Pred. No.:                             | 4..2e-59   | Length:       | 830 |
| Score:                                 | 973.50   | Matches:      | 220 |
| Percent Similarity:                    | 56.97%   | Conservative: | 115 |
| Best Local Similarity:                 | 37.41%   | Mismatches:   | 236 |
| Query Match:                           | 22.88%   | Indels:       | 17  |
| DB:                                    | 1  | Gaps:         | 7   |
| AF133659 (1-2345) x HMT1_SCHPO (1-830) |  |               |     |
| QY 370                                 | ATGCTTTCTTATGTGTGCCCC--AAAGC--AGCCAGATCTACAGACTAGAGTTGCC       | 423           |     |
| DB 249                                 | LeuLeuProHyLeuTrProThrHyAspTyArgLeuGlnPheGlnIlePheIleCys       | 268           |     |
| QY 424                                 | ATTTCGGTGGATTTTGGGTGGTGCAGAAAGCCATGAATTTGGTCCCTTCATGTTT        | 483           |     |
| DB 269                                 | IleValLeuLeuPheLeuGly-----ArgAlaValAsnIleLeuAlaProLysGlnLeu    | 286           |     |
| QY 484                                 | AAATATCTGTAGACACGCTCAACAGATGTGGGAAACATGCTGAACCTGAGTATGCA       | 543           |     |
| DB 287                                 | GlyValLeuThrGlnLysLeuThrLysHisSer---GlyLysIleProTrpSerAsp---   | 304           |     |
| QY 544                                 | CCAAATACAGTTGCACACCATGACGACACAGCATTTGGATTGGCTATGGTATCAAGAGCT   | 603           |     |
| DB 305                                 | -----ValIleLeuPheValIleLeuThrArgPheLeuGlnIly                   | 316           |     |
| QY 604                                 | GGAGCTGCTTTTTTAAAGAGTTGCAGAAATGCAAGTATTTGGACAGGTAGCCAGAAATCA   | 663           |     |
| DB 317                                 | AsnMetGlyValIleGlySerLeuAspGserPheLeuTrpValProValSerGlnTyAla   | 336           |     |
| QY 664                                 | ATCCGAGAAATAGCAGAAAAATGCTTTTCTCATCTTTCACACACCTGGATCTGGGTTTTCAC | 723           |     |
| DB 337                                 | TyrTrpGlnAlaIleSerThrLysAlaLeuArgHisValLeuAsnLeuSerTyArgPheHis | 356           |     |
| QY 724                                 | CTGAGCAGACAGACGGAGCTTTATCTTAAGCTATTGACAGAGAAACAAGGGATACGT      | 783           |     |
| DB 357                                 | LeuAsnLysArgAlaGlyValLeuThrAlaLeuThrLysGlySer---SerLeuAsn      | 375           |     |
| QY 784                                 | TTTGCTCGAGTGTGGTGGTATTTATCTTTCCCATCATGTGTTGAAGTAGTGGTTTGC      | 843           |     |
| DB 376                                 | ThrPheAlaGlnGlnValValPheGlnIleGlyProValLeuLeuAspLeuIlyAla      | 395           |     |
| QY 844                                 | AGTGGTGTGTTTGATTAACAATGCGGCGCCAGTTGGCTTGGTAACTTGGAAACACTT      | 903           |     |
| DB 396                                 | MetValTyrrPhePheIleLysPheAspIleTyrrPheThrLeuIleValLeuIleMetThr | 415           |     |
| QY 904                                 | GGTACATACACAGCATTCACAGTTGCAGTCACACGGTGGAGAACTAGATTGAATAGAA     | 963           |     |
| DB 416                                 | LeuGlySerTyrrGlyValThrValLysIleTnSerTrpArgThrGlnAlaArgLys      | 435           |     |
| QY 964                                 | ATGAAACAAGCAGATTAATGATGCAAGGTAATGCTGCTATAGACTACGTGTAATTATGAA   | 1022          |     |

Db 436 MetValIenSerTrpAglUSeTrAlaValAlnIenSapAlaIleMetAsnPhgGlu 455  
 QY 1024 ACTGTGAATTTTAAATGAAGAATATGACACAGATATGATGATTTTGAAG 1083  
 Db 456 ThrValIysAsnPhenAspAlaAspAspPheGluIenGluArgIylGlyHlaValAla 475  
 QY 1084 ACGTATGAGACTGCTTATGAAAAGTACTTACTCTGCTTATGCTGAACTTGTCAA 1143  
 Db 476 IleTrpIysGlnGluArgIylValIleuPheSerIeuSnpheAsnIleValGln 495  
 QY 1144 AGTGTCTTTTACGTGCGTTTAAAGCTTATGCGTCGCGACAGCAGGAATTGTG 1203  
 Db 496 GlyGlyIlePheTrpPheSerIeuAlaIleAlaCysIeuSeraIAlaTrpValThr 515  
 QY 1204 GCAAGTACCTTACTGTGAGATCTAGTATGAGTGAATGAGCTTTTCACTTCA 1263  
 Db 516 PheGlyPheAsnThrValIglAspPheValIleuIeuThrTrpMetIleGlnIeuGln 535  
 QY 1264 TTACCTCTGAACCTTTGCGGAACCTGTATATGAGAGACAGCAGCAGCTATGATG 1323  
 Db 536 GlnProIeuSnphePheGlyThrIeuTrpArgSerIeuGlnAsnSerIleIleAspThr 555  
 QY 1324 AACACCTTTTACTCTACTCAAGTATGACACCCAAATTAAGCAAGTATGATCT 1383  
 Db 556 GluArgIeuGlnIlePheGluIylUylProThrValValGluIylProAsnAlaPro 575  
 QY 1384 CCCCTTGATCAGACACACAGACAGTACCGCTTGTATATGATTTTGAATAC 1443  
 Db 576 AspIeuIysValThr-----GlnGlyIylValIlePheSerHlaValSerPheAlaTrp 593  
 QY 1444 ATTGAGGCGCAAGAAATCTTATGGAATATCTTGAAGTCCCTGACAGGAAGAAAGT 1503  
 Db 594 AspProArgIylProValIeuSerAspIleAsnIleValAlaGlnProGlyIylValIle 613  
 QY 1504 GCCATTGTGAGATGATGCTGACGAGAAAAGCAGATAGAGAGCTATTTTCCTTC 1563  
 Db 614 AlaIeuValIglUylSerGlyIylGlyIylSerThrIleMetArgIleIleuIeuArgPhe 633  
 QY 1564 TATGAGCTCAAAAGGTAGATCTTATCTGCTGCTGCAATATATACAGATGAGCTG 1623  
 Db 634 PheAspValIenSerGlySerIleThrIleAspAspIleAsnIleValAlaThrIleu 653  
 QY 1624 GAAAGCTTGGAGGAGGAGGAGGAGTGTACTGAGATGCTGCTTCCATATATACT 1683  
 Db 654 SerSerIeuArgSerIleGlyValIleValProGlnAspSerThrIeuPheAsnAspThr 673  
 QY 1684 ATTATATACACCTCTTATATGAAACATCAGTCTTCACTGAGGAAGTATGCGATG 1743  
 Db 674 IleIeuTrpAsnIleIylTrpAlaIylProSerIleThrAsnGluIylIleTrpAla 693  
 QY 1744 GCAAAATTTAGCTGACTTATGATGATGCAATCTTGAATGCGACATGATATGACCCAA 1803  
 Db 694 AlaIylAlaIleGlnIleHlaSapArgIleIeuSnpheProAspGlyTrpAsnSerArg 713  
 QY 1804 GTAAGGGAACGAGACTCAAGCTTTGAGAGAGAAAAGCAAGATGAGCAATGCAAGA 1863  
 Db 714 ValGlyIylUylArgIylIeuIylSerGlyIylGlyIylUylSerIleValAlaValArg 733  
 QY 1864 GCCATTTTGAAGACCCCAAGTATCTCTATGATGAGTACTTATGCTGATTCG 1923  
 Db 734 AlaIleIeuIylAspProSerIleIleIleuIeuAspGluIleThrSerIleAlaIeuAspThr 753  
 QY 1924 ATTACTGAAGACTATTTCTGTGCGATGAAGATGCGCAACAGCAAGACTTAT 1983  
 Db 754 AsnThrIylUylArgIylGlnIleGlnAlaIleuSnpArgIleuIleAspGlyArgThrAlaIle 773  
 QY 1984 TTCAATGACACAGATTTGTCAACAGTGTATGACAGATGAATCATGCTTGGATCAG 2043  
 Db 774 ValIleHlaIleAspArgIeuSerThrIleThrAsnAlaAspIleuIleuIylSerAsn 793  
 QY 2044 GGTAAAGTACCCGAAAGCTGTATCCACCATGCTTGTCTGTAACCTCATATGATCTAT 2103  
 Db 794 GlyArgIleValIglUylThrIleHlaSglUylIleuIleIylSapArgIylGlyAlaTrp 813

QY 2104 TCAGAAATGTGCATACACAGACC 2127  
 Db 814 LysIleuSerTrpPheGlnGlnAla 821  
 RESULT 7  
 YFIC\_BACSU  
 ID YFIC\_BACSU STANDARD; PRT; 604 AA.  
 AC P54719;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical ABC transporter ATP-binding protein yf1c.  
 GN yf1c.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.  
 OX NCBI\_Taxid:1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96262713; PubMed=8704981;  
 RA Yamamoto H., Uchiyama S., Fajar A.N., Ogasawara N., Sekiguchi J.;  
 RT "Determination of a 12 kb nucleotide sequence around the 76 degrees  
 RL Microbiology 142:1417-1421 (1996).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertolo M.G., Bessieres P., Bolocin A., Borchert S.,  
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch J., Caldwell B., Camargo V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entlin K.D., Errington J., Fabret C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,  
 RA Gilm S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Hatech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,  
 RA Jorje B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber V., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moesl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portecelle D., Porwollik S., Prescott A.M.,  
 RA Priescan E., Pujic P., Putrelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan B., Schleich S., Schroeder R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seir S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Toseco V., Uchiyama S., Vandenbol M., Vannier F., Vassaretli A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Welzenegger T.,  
 RA Winters P., Wiput A., Yamamoto H., Yamane K., Yatsunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RL subtilis".  
 RL Nature 390:249-256 (1997).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MSBA SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; D50543; BAA09107.1; -  
 CC EMBL; Z59108; CAB12651.1; -

```

DR HSSP; P13569; 1MBD.
DR Subtilisin; BG1850; Yf1C.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transport.
DR InterPro; IPR01140; ABC_transport.
DR Pfam; PF00664; ABC_membrane; 1.
DR ProDom; PD00006; ABC_transport; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR Hypothetical protein; ATP-binding; Transport; Transmembrane;
KW Complete proteome.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 86 106 POTENTIAL.
FT TRANSMEM 172 192 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
FT TRANSMEM 510 530 POTENTIAL.
FT NP_BIND 399 406 ATP (POTENTIAL).
SQ SEQUENCE 604 AA; 67396 MW; 1998E469A2E6133 CRC64;

Alignment Scores:
Pred. No.: 1,12e-38 Length: 604
Score: 674.50 Matches: 190
Percent Similarity: 49.52% Conservative: 120
Best Local Similarity: 30.35% Mismatches: 257
Query Match: 15.86% Indels: 59
DB: 1 Gaps: 16

AF133659 (1-2345) x Yf1C_BACSV (1-604)
QY 310 CTCACACAGACCCAAAGAGGTTAAAGATGTGTAAGTCTCGAATAATCATAAAGCA 369
DB 14 LeuProIleAspLysLysGluGluGlyAlaLysLysArgAlaLysAlaLysAspThrLysGly 33
QY 370 ATGCTTTTGTATGTGTGCC-----AAAGACAGCCAGATCTCGAGCTTGAAGTTGCC 423
DB 34 ThrLeuArgArgLysIleTrpSerTyrLeuAlaGluArgLysGlyLeuLeuIleLeuValMet 53
QY 424 ATTTCGCTG-----GGATTTTGGGGTGGTGCAGCAAGGCCATGAATATT 465
DB 54 LeuMetValIleLysSerAlaIlePheGlyLeuLeuGly----- 66
QY 466 GTGTTTCCCTTCATGTTAAATATGCTGACAGCCCTC-----AACGAGATGTGG 516
DB 67 -----ProPheValIleGlyLysAlaIleAspHisPheIleValGlyLysThrValSer 84
QY 517 GGAACATGCTGAACCTGAGTGCACCAAAATACAGTTGCAACCATGGCAACAGCAGTT 576
DB 85 GlyLeuIle-----ProValLeuLeuLeuLeuAlaIle 96
QY 577 CTGATTTGGCTATGTTGATCAGAGCTGAGTGTCTTTTAAAGAGTTGGAATGCA 636
DB 97 TyrIleIleGlnSerLeuSer-----LeuTrpPhe-----GlnAsnTyr 109
QY 637 GTATTTCGCAAGTACCCAGATTCATCCGAGAAATAGCCAAATGCTTTCTTCAT 696
DB 110 TrpMetIleThrIleSerGlnGlyThrValPheArgMetArgSerGlnLeuPheThrHis 129
QY 697 CTTCACAACTGGATCTGGGCTTTTCACTGAGCAGACAGACGGAAGCTTATCTAAGCT 756
DB 130 LeuHisGluLeuProIleProPhePheAspLysGlnArgHisGlyLysMetSerArg 149
QY 757 ATTGACAGAGCAAGAGGGTATCAGTTTGTCTGAGTGGCTTGGTATTAATCTTCTT 816
DB 150 ValThrAsnAspIleGluAsnValSerSerThrLeuAsnThrSerValIleGlnIleLeu 169
QY 817 CCCATCATGTTTGAAGTATGCTTCAGTGGTGTGTTTGTATTAATAATGCGGTCCAG 876
DB 170 SerSerValIleThrPheValGlyThrIleAlaValMetLeuTyr-----Met 185
QY 877 TTTCGTTTGGTAACCTTGGGAACACTGTGTACATACACAGATTTCACAGTTCAGTCA 936
DB 186 SerProLeuLeuThrLeuIleThrLeu---ThrIleIleProValMetAlaIleSerLeu 204

```

```

QY 937 CGGTGAGAACTAGA-----TTTAGAATAGAAATGAACAAAGCAGATAAT 981
DB 205 LysTrpIleThrAsnArgThrGlyLysLeuPheLys---GluGlnGlnLysAsnLeuGly 223
QY 982 GATCAGAGTAATGCTGCTATAGACTCACTGCTGAATATGAAACTGGAAGTATTTAAT 1041
DB 224 AspLeuasnGlyTyrIleGluGlnSerValSerGlyAlaLysValIleLysAlaTyrSer 243
QY 1042 AATGAAGATATGAAGACAGAGATATGATGATTTTGAAGAGTATGAGCTGCTCA 1101
DB 244 ArgGluLys-----GlnIleThrAlaGluPheLeuGlu-----LysAsnAlaIle 258
QY 1102 TTGAAAGTACTCTACTCTGCTATGCTGAACCTTTGGT-----CAAAAGT 1146
DB 259 LeuLysThrSerGlyPheTrpAlaGlnThrIleSerGlyPheIleProLysValMetAsn 278
QY 1147 GCTATTTTCACTGTCGGTTTAAACCTATTAATGCTGCTCCGCACTCAGGAATTTGCA 1206
DB 279 SerLeuAsnLeuSerPheThrMetIleAlaIleIleGlyLysLeuPheAlaLeuLys 298
QY 1207 GGTACCTTACTGTTTGGAGATCTATGTAATGCTGAATGAGCTTTTTCAGCTTCAATTA 1266
DB 299 GlyTrpIleSerIleGlySerIleValValPheAlaGluTyrSerArgGlnPheThrArg 318
QY 1267 CCCCTGAACCTTCTGGGAACTGATATAGAGAGACTAGACACAGCACTCATGATATGAAAC 1326
DB 319 ProLeuAsnAspLeuAlaAsnGlnPheAsnThrMetLeuSerAlaIleAlaGlyAlaGlu 338
QY 1337 ACCTGTTTACTCTATCTCAAGTATGACACCCAAATTAAGCAAAAGTATGCTTCCC 1386
DB 339 ArgValPheAspValIleAspGluLysGluGlnArgGlnAspGlyLysAsnAlaValHis 358
QY 1387 CTTCAGATCACACACAGACAGCTACCGTGCTTTGATATGTCATTTTGAATATACATT 1446
DB 359 GlnProIle-----GlnThrGlySerIleGluPheArgAspValSerPheGlyTyrAsp 376
QY 1447 GAGGCGCAAGAGTCTTATGTAATATCCTTGAAGTCCCTCGAGGAAAGAAAGTGCC 1506
DB 377 LysGlyGlnGlnThrLeuLysHisIleLeuGlnPheThrValProAlaGlyGlnSerIleAla 396
QY 1507 ATTTAGAGGTAGTGGTCAAGGAAAGCAAAATAGTACGCTATTAATTTGCTTCTAT 1566
DB 397 PheValGlyProThrGlyAlaGlyLysThrThrValThrAsnLeuLeuAlaArgPheTyr 416
QY 1567 GAGCTCAAAAGGTAGATTTATCTTGCTGCTGCAAAATATACAAAGTACGCTGGA 1626
DB 417 GluProAsnAspGlyLysIleLeuIleAspGlyThrAspIleLysThrLeuThrArgAla 436
QY 1627 AGCCTTCGAGGAGGAGTGGAGTGTACTCAGAGATGCTGCTTCATATATCTAAT 1686
DB 437 SerLeuArgLysAsnMetGlyPheValLeuGlnHisPheLeuPheGlnGlyThrIle 456
QY 1687 TATTACAACTTTATATAGAAATCATGCTTCACTCAGAGAGTGTATGACGTGCA 1746
DB 457 ArgGluAsnIleArgTyrGlyArgLeuAspAlaSerGlnGluValIleAlaIleAla 476
QY 1747 AATTTAGCTGATCTCATGATGCAATTTCTTGATGCAATGCAATGATATGACACCAAGTA 1806
DB 477 LysThrAlaAsnAlaHisSerPheIleGluArgLeuProLysGlyTyrAspThrValLeu 496
QY 1807 GGGGACAGAGACTCAAGCTTTTCAGAGAGAGAAAGCAAAAGTATGCAATGCAAGACC 1866
DB 497 ThrGlnAsnGlySerGlyIleSerGlnGlyLysGlnLeuIleSerIleAlaIleArgAla 516
QY 1867 ATTTTGAAGACCCCGCATACTATGATGAAAGTCACTTATGATGATTCGATT 1926
DB 517 ValLeuAlaAspProValLeuLeuIleAspGluAlaThrSerAsnIleAspThrVal 536
QY 1927 ACTGAAGACTATTTCTTGGTGCATGAAAGATGTGGTCAAAACACAGAACTTATTTTC 1986
DB 537 ThrGluValAsnIleGlnGluAlaLeuAlaArgLeuMetGluGlyArgThrSerValIle 556

```

QY 1987 ATTCGACAGCATGTCACAGCTGTTGATGACATGAATCATTTGTCATCAGGT 2046  
|||||  
DB 557 ILeAlahIsaGlueuAaenthrILeIlgAargAlaAspInILeValLeuLysAsnGly 576  
|||||  
QY 2047 AAGGTACCGCAACGTGTACCATCATGTTGCTTCTTAACCTCATAGATATATTC 2106  
|||||  
DB 577 GluMetILeGluLysGlySerHisAspGluLeuIle---ArgInLysGlyPheTyrSer 595  
|||||  
QY 2107 GAATGTGCATACACAG 2124  
|||||  
DB 596 AspleuTyrGluSerGln 601  
|||||  
RESULT 8  
HLI2\_ECOLI  
ID HLI2\_ECOLI STANDARD; PRT; 707 AA.  
AC P10089;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemolysin secretion ATP-binding protein, chromosomal.  
GN HLYB.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OX NCBI\_taxid=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J96 / Serotype O4;  
RX MEDLINE=85234404; Pubmed=3891743;  
RA Felmei T., Pellet S., Welch R.A.;  
RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin";  
RL J. Bacteriol. 163:94-105(1985).  
CC -1- FUNCTION: INVOLVED IN THE EXPORT OF HEMOLYSIN A.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. HLYB SUBFAMILY.  
CC  
CC This SMS-S-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sb-sib.ch).  
CC  
CC EMBL: M10133; AAA23976.1; -.  
DR PIR; B24433; LEECB.  
DR MEROPS; C39; UNK; -.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR001140; ABC\_transporter\_TM.  
DR InterPro; IPR005074; Peptidase\_C39.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR Pfam; PF00664; ABC\_membrane; 1.  
DR Pfam; PF03412; Peptidase\_C39; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
DR  
KM Hemolysis; Transport; ATP-binding; Transmembrane.  
FT TRANSMEM 158 179  
FT TRANSMEM 187 204  
FT TRANSMEM 269 289  
FT TRANSMEM 293 311  
FT TRANSMEM 360 377  
FT TRANSMEM 387 403  
FT TRANSMEM 407 425  
FT TRANSMEM 502 509  
SQ SEQUENCE 707 AA; 79463 MW; 21009CB45E59437E CRC64;  
Alignment Scores:  
Pred. No.: 1,34e-38 Length: 707  
Score: 673.50 Matches: 176  
Percent Similarity: 51.69% Conservative: 115

Best Local Similarity: 31.26% Mismatches: 239  
Query Match: 15.83% Indels: 33  
DB: 1 Gaps: 11  
AF133659 (1-2345) x HLI2\_ECOLI (1-707)  
QY 463 ATTCGTGTCCTTCATGTTAAATATGCTGTAGACAGCTCAACAGATGTCGGA--- 519  
|||||  
DB 169 LeuILeThrProLeuPhePheGlnValValMetAspLysValLeuValHisArgGlyPhe 188  
|||||  
QY 520 AACATGCTGAACCTGATGATGACCAAAATPACGTGGCAAGGCAACAGAGCTTCG 579  
|||||  
DB 189 SerThrLeuAsn-----ValILeThrValAlaLeuSerValVal 201  
|||||  
QY 580 ATTCGATATGCTATCAAGAGCTGAGCTGCTTTTAAACGAATTCGAATGACGTA 639  
|||||  
DB 202 ValValPheGluLeuLeuLeuSerGly-----LeuArgThrTyrIle 215  
|||||  
QY 640 TTTCGAAGTACCCAGAAATTCATCCAGAAATAGCCAAAATGCTTTTCATCTT 699  
|||||  
DB 216 PheAlaHisSerThrSerArgIleAspValGluLeuGlyValAlaLysLeuPheArgHisLeu 235  
|||||  
QY 700 CACAACCTGATGCGGTTTACCTGACAGACAGACAGCGGAGCTTATTAAGCTATT 759  
|||||  
DB 236 LeuAlaLeuProILeSerThrPheGluSerArgArgValGlyAspThrValAlaArgVal 255  
|||||  
QY 760 GACAGAGAAACAAGGGTATCAGTTTGTCTGAGGCTTTGGTATTAATCTTCC 819  
|||||  
DB 256 ArgGluLeuAspGlnILeArgAsnPheLeuThrGlyGlnAlaLeuThrSerValLeuAsp 275  
|||||  
QY 820 ATCATGTTGAAGTATGCTGTGTCAGTGTGTTTGTATTAACAATCGGTCCAGTTT 879  
|||||  
DB 276 LeuLeuPheSerPheILePhePheAlaValMetTyrTyr-----SerProLysLeu 293  
|||||  
QY 880 GCTTGTGAACCTTGGAACCTGCTGATACATACAGACGATTCACAGTGCAGACACA--- 936  
|||||  
DB 294 ThrLeuValILeLeuPheSerLeuProCysTyrAlaAlaThrSerValPheILeSerPro 313  
|||||  
QY 937 -----CGGTGGAACCTGATTAAGATTAAGAAATGAACAAAGACAGATATGACAGT 990  
|||||  
DB 314 ILeLeuArgArgArgLeuAspAspLysPheSerArgAsn---AlaAspAsnGln----- 330  
|||||  
QY 991 AATGCTGATATGACTCACTGCTGAAATTAAGAACTGTAAGTATTTAATAAGAAAGA 1050  
|||||  
DB 331 SerPheLeuValGluSerValThrAlaIleAsnThrILeLysAlaMetAlaValSerPro 350  
|||||  
QY 1051 TATGAAGCACAGAGATATGATGATTTTGAAGACGTATGACGCTTCATGAAGAAAGT 1110  
|||||  
DB 351 GluMetThrAsnILeThrAspLysGlnLeuAlaGlyTyrValAlaIleAspPheLysVal 370  
|||||  
QY 1111 ACCTCACTGTGGCTATGCGTGAACCTTGTGCAAGGCTATTTTCAGTGCAGTTTAACA 1170  
|||||  
DB 371 Thr-----ValLeuAlaThrILeGlyGlnGlnGlyILeGlnLeuILeGlnLysThr 387  
|||||  
QY 1171 GCTATATATGATG-----CTCGCACTCAAGGAATGTGGCAGTACCTTACTGTT 1221  
|||||  
DB 388 ValMetILeILeAsnLeuThrPheGluAlaHisLeuValILeSerGlyLysLeuSerIle 407  
|||||  
QY 1222 GGAAGATCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1281  
|||||  
DB 408 GlyGlnLeuILeAlaPheAsnMetLeuAlaGlyGlnILeValAlaIleProValILeArgLeu 427  
|||||  
QY 1282 GGAAGCTATATATGAGAGCTAGCAAGACATCATGATATGAGACCTTGTATTACTGA 1341  
|||||  
DB 428 AlaGlnILeThrPheGlnPheGlnGlnValGlyILeSerValThrArgLeuGlyAspVal 447  
|||||  
QY 1342 CTCAGATGAGACCCCAATTAAGACAAAGTATGATGATGATGATGATGATGATGATG 1401  
|||||  
DB 448 LeuAsnSerProThrLeuSerTyrHisGlyLysLeuAlaIlePro-----Glu 463  
|||||  
QY 1402 CAGACAGCTACCGTGCCTTTGATATATGATGATGATGATGATGATGATGATGATGATG 1458  
|||||  
DB 464 ILeAsnGlyAspILeThrPheArgAsnILeArgPheArgTyrLysProAspSerProVal 483  
|||||

|            |  |   |              |
|------------|--|---|--------------|
| Oy         | 1459   | GTCCTACTGGAATATACCTTTGAAGTCCCTCGCAGGAAAGAAAGTGCCATTGTAGAGAGT    | 1518         |
| Db         | 484  | IleuenaapaaNilleaNIleuSerIleYelngInclYgluaalIlelYllleValaIGlyar | 503          |
| Oy         | 1519   | AGTGGGTGAGGAAAAGACACAATATGAGGCTATTAATTCCTTCATGAGGCTCAAGAAG      | 1578         |
| Db         | 504  | SergIysergIylYserThrIeuThrIeuIlleGlHarghethrIIeProGluN          | 523          |
| Oy         | 1579   | GGTAGCATTTATCTTGCTGGTCGCAAAATATATCAAGATGTGAGCTGGAAAGCCTTGGAGG   | 1638         |
| Db         | 524  | GlylnalValleuIlleaPglyHIsasPleaalaIleualaasProaantPreuuarGar    | 543          |
| Oy         | 1639   | GCAGGGGAGGTGATACCTCCAGGATGCTGTGCTCTCCATTAATCTATTTATTAACAATC     | 1698         |
| Db         | 544  | GlnValIGlyvalIValleuglnasPasnaVallleuleuashrgserIlelaaspnIle    | 563          |
| Oy         | 1699   | TTATATGGAACATCAGTGCTTCACCTGAGAAAGTATGCAAGTGGCAAAATTAAGCTGA      | 1758         |
| Db         | 564  | SerIleualaIeaenProGlyMeSerValIGlutlyValIIetYalaalIalsyeunlaGly  | 583          |
| Oy         | 1759   | CTTCATATGCAATTTCTTCGAATGCCACATGATATGACACCACCAAGTAGGAGGACAGAGA   | 1818         |
| Db         | 584  | AlahIsaSPheIIeSerGIuearIGluIGlyTarasnThrlleValIGlyluIngIay      | 603          |
| Oy         | 1819   | CTCAAGCTTTGAGGAGGAGAAACAAAGAGTGGCAATGCAAGACGCTTTTGAAGAGC        | 1878         |
| Db         | 604  | AlaaglyleuSerGIyglYalnIGlnarIlealllealIeArgralaIeualaleuNshen   | 623          |
| Oy         | 1879   | CCCCAGTCATACTCTATGATGAAGACTACTTCATCGTTAGATTGCAATTACTGAGAGACT    | 1938         |
| Db         | 624  | ProlYslIleuIlePheaSPgluaIathSerAlaleuasprrglusergIuhIval        | 643          |
| Oy         | 1939   | ATTCTTGTCGCATGAAGAGATGTGCTAAACACAGAACTTATTTTCAATTCACACAGA       | 1998         |
| Db         | 644  | IleMetAysaMeethIslyslIeCySlysgIyarGrhvaalIlellelealahaIsar      | 663          |
| Oy         | 1999   | TTCGTAACAGGTGGTGGATGACGATGAATCATTTGCTGGATCAGGGTAGGTACCCAA       | 2058         |
| Db         | 664  | LeuSerThrVallysaaNaIaAsparGIlelleValaMecIuluSelyslYsrIleValaGu  | 683          |
| Oy         | 2059   | CGTGATCCACCAACCATGGTTGCTGTCTAACCTCCATCAATGATATTCAGAAAATGGGCAT   | 2118         |
| Db         | 684  | GlnGlylsHisIysGIuLeuLeuSerGIuProGluSerIeuYrserTYreutyrGln       | 703          |
| Oy         | 2119   | AACACAGAGC 2127   |              |
| Db         | 704  | LeuGINser 706   |              |
| <hr/>      |  |   |              |
| RESULT 9   |  |   |              |
| HLXB_ECOLI |  |   |              |
| ID         | _HLXB_ECOLI  | STANDARD;   | PRT; 707 AA. |
| AC         | P08716;  |   |              |
| DT         | 01-JAN-1988 (Rel. 06, Created)                                       |   |              |
| DT         | 01-JAN-1988 (Rel. 06, last sequence update)                          |   |              |
| DT         | 16-OCT-2001 (Rel. 40, last annotation update)                        |   |              |
| DE         | Hemolysin secretion ATP-binding protein, plasmid.                    |   |              |
| CN         | HLXB.  |   |              |
| OS         | Escherichia coli.  |   |              |
| OC         | Plasmid Inc12 pHyV152.   |   |              |
| OC         | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;     |   |              |
| OC         | Escherichia.   |   |              |
| OX         | NCBI_TaxId=562;<br>[1]   |   |              |
| RN         | SEQUENCE FROM N.A.   |   |              |
| RP         | Hess J., Wells W., Vogel M., Goebel W.;                              |   |              |
| RA         | "Nucleotide sequence of a plasmid-encoded hemolysin determinant and  |   |              |
| RT         | its comparison with a corresponding chromosomal hemolysin sequence." |   |              |
| RL         | FEMS Microbiol. Lett. 34:1-11(1986).                                 |   |              |
| RN         | [2]  |   |              |
| RP         | TOPOLOGY.  |   |              |
| KX         | MEDLINE=92204133; PubMed=1552901;<br>Gentschev I., Goebel W.;        |   |              |

```

RT      "Topological and functional studies on HlyB of Escherichia coli.";
RL      Mol. Gen. Genet. 232:40-46(1992).
RN      [3]
RP      TOPOLOGY.
RX      MEDLINE=91113653; PubMed=1994034;
RA      Wang R.C., Seor S.Y., Blight M., Pratt J.M., Broome-Smith J.K.,
RA      Holland I.B.;
RT      "Analysis of the membrane organisation of an Escherichia coli protein
RT      translocator, HlyB, a member of a large family of prokaryote and
RT      eukaryote surface transport proteins.";
RL      J. Mol. Biol. 217:441-454(1991).
CC      -1- FUNCTION: INVOLVED IN THE EXPORT OF HEMOLYSIN A.
CC      -1- SUBUNIT: HOMODIMER.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC      -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. HlyB SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL, M14107; AAA98234.1; -.
DR      PIR, S10057; S10057.
DR      MEROPS, C39.UNW; -.
DR      InterPro: IPR003593; AAA_ATPase.
DR      InterPro: IPR003439; ABC_Transport.
DR      InterPro: IPR001140; ABCTransportTM.
DR      InterPro: IPR005074; Peptidase_C39.
DR      Pfam, PF00005; ABC_tran; 1.
DR      Pfam, PF00664; ABC_membrane; 1.
DR      Prodom, PD00006; ABC_transportr; 1.
DR      SMART, SM00382; AAA; 1.
DR      PROSITE, PS00211; ABC_TRANSPORTER; 1.
KW      Hemolysis; transport; ATP-binding; Transmembrane; Plasmid.
FT      DOMAIN 1 37
FT      TRANSMEM 38 51
FT      DOMAIN 52 57
FT      TRANSMEM 58 78
FT      DOMAIN 79 157
FT      TRANSMEM 158 179
FT      DOMAIN 180 190
FT      TRANSMEM 191 211
FT      DOMAIN 212 268
FT      TRANSMEM 269 289
FT      DOMAIN 290 294
FT      TRANSMEM 295 315
FT      DOMAIN 316 387
FT      TRANSMEM 388 408
FT      DOMAIN 409 411
FT      TRANSMEM 412 432
FT      DOMAIN 433 707
FT      NP BIND 502 509
SQ      SEQUENCE 707 AA; 79672 MW; 412A3BB64A3CFFBA CRC64;

Alignment Scores:
Pred. No.: 1.83e-38
Score: 671.50
Percent Similarity: 51.69%
Best Local Similarity: 31.08%
Query Match: 15.79%
DB: 1 Gaps: 11

AF133659 (1-2345) x HLYB_ECOLI (1-707)
QY      463 ATTTGGTTCCCTTCATGTTAAATATGCTGAGACGCTCAACAGATGCGGA--- 519
DB      169 LeuIeIhrrProLeuPherheGlnIaValaMetabryValaIeValaHieIargIyhe 188
QY      530 AACATGCTGAACCTTGATGATGACCAACCAATATACAGTTGACACCATGCAACAGCACTTCTG 579

```

```

Db      189 SerThrLeuAen-----ValIleThrValAlaLeuSerVal 201
QY      580 ATTGGCTATGGTATCAAGAGCTGGAGCTGTTTAAAGAAATGCAAGTGA 639
Db      202 ValValPheGluIleIleLeuSerGly-----LeuArgThrTyrIle 215
QY      640 TTGGCAAGTAGGCCCAAAATTCGAATCCGAAGAAATAGCCAAAATGCTTTCATCTT 699
Db      216 PheAlaHisSerThrSerArgIleAspValGluLeuGlyAlaValSerPheArgHisLeu 235
QY      700 CACAACCTCGATCTGGGTTTACCTGAGCGACAGACAGCGAGCTTATCTTAAGCTATT 759
Db      236 LeuAlaLeuPheProIleSerTyrPheGluSerArgValGlyAspThrValAlaArgVal 255
QY      760 GACAGAGAAACAAGGGGTATGATCTTGTCTGAGTGCTTGATTTAATCTTCCCTCC 819
Db      256 ArgGluLeuAspGlnIleArgAsnPheLeuThrGlyAlaLeuThrSerValAsp 275
QY      820 ATCATGTTTGAAGTATGATGCTTGTGACGGTGTGTTGATTAACAAATGGGCTCCAGTTT 879
Db      276 LeuLeuPheSerIleIlePhePheAlaValMetTrpTyrTyr-----SerProGlyLeu 293
QY      880 GCTTGGTAACCTTGGAACTTGGTGAACATACACAGCATTCACAGTTGCAGTACA--- 936
Db      294 ThrLeuValIleLeuPheSerLeuProCysTyrAlaIleTrpSerValPheIleSerPro 313
QY      937 -----CGGTGAGACATGATTTGAATGAATGAACAAGACAAATGATGACAGT 990
Db      314 IleLeuArgIleArgIleAspAspIlePheSerArgAsn--AlaAspAsnGln----- 330
QY      991 AATCTGCTATAGATCACTGCTGCTGAATTAAGAACTGGAAGTATTTAATGAAGA 1050
Db      331 SerPheLeuValGluSerValThrAlaIleAsnThrIleGlyAlaMetAlaValSerPro 350
QY      1051 TATGAAGACAGAGATATGATGATTTTGAAGACGATGAGACTGCTTCAATGAAGAAGT 1110
Db      351 GlnMetThrAsnIleTrpAspIleGluLeuAlaGlyTyrValAlaAlaGlyPheGlyVal 370
QY      1111 ACCCTTACTGCTGCTGATGCTGAACCTTGGTCAAGGCTGATTTTCAGTTCGGTTTACA 1170
Db      371 Thr-----ValLeuAlaThrIleGlyGlnGlnGlyIleGlnLeuIleGlnGlyLeuThr 387
QY      1171 GCTATAATATGCTG-----CTGCCAGTCAAGGGAATTTGGCAGGATCCCTTACTGTT 1221
Db      388 ValMetIleIleAsnLeuThrPheGlyAlaHisLeuValIleSerGlyAspLeuSerIle 407
QY      1222 GGAGATTTAGTATGATGATGATGATGCTGCTTTTTCAGCTTTCATTAACCTTCACTG 1281
Db      408 GlyGlnLeuIleAlaPheAsnMetLeuAlaGlyGlnIleValAlaProValIleArgLeu 427
QY      1282 GGAAGCTATATAGAGACTAGACATGACAGCACTCATATATGAAACACTTGTACTCTA 1341
Db      428 AlaglnIleTrpGlnAspPheGlnGlnValGlyIleSerValThrArgGlyAspVal 447
QY      1342 CTCAGGTAGACACCAATTAAGACAAAGTGAATGATGCTCCCTTCAGATCACACCA 1401
Db      448 LeuAsnSerProThrGlu-----SerTyrHisGlyGlySerLeuThrLeuProGlu 463
QY      1402 CAGACAGCTACCTGGCTTGTATGATGTCATTTGAATAC--ATTGAGGGCCAGAAA 1458
Db      464 IleAsnGlyAspIleThrPheArgAsnIleArgPheArgTyrLeuProAspSerProVal 483
QY      1459 GTCCTTAGTGAATATCTTTGAAGTCCCTGCCAGAAAGAAAGGCGCATTTAGAGAGT 1518
Db      484 IleLeuAspAsnIleAsnLeuSerIleGlyGlnGlnGlyValIleGlyIleValGlyArg 503
QY      1519 AGTGGGTCAAGGAAACACAAATAGTAGAGCTATTATTTGGCTTCTTGAAGCCCAAG 1578
Db      504 SerGlySerGlyIleGlySerThrLeuThrLeuLeuIleGlnArgPheTyrIleProGlnAsn 523
QY      1579 GGTAGCATTTATCTTCTGCTGCTCAAAATATATCAAGATGTGAGCTTGAAGACCTTGGAGG 1638

```

```

Db      524 GlyGlnValLeuIleAspGlyHisAspLeuAlaLeuAlaAspProAsnTrpLeuArg 543
QY      1639 GCAGTGGAGTGGTGAACCTCAGATAGCTGCTCTTCATATATCTATTATTAACAACCTC 1698
Db      544 GlnValGlyValValLeuGlnAspAsnValLeuLeuAsnArgSerIleIleAspAsnIle 563
QY      1699 TTATATGAAACATCACTAGTCTTCACTGAGGAAGTATGACAGTGGCAAAATTAAGCTGA 1758
Db      564 SerLeuAlaAsnProGlyMetSerValGluValIleTyrAlaAlaValSerLeuAlaGly 583
QY      1759 CTTTCAATGATGCAATTTCTTCAATGCCACATGATATGACACCCCAAGTAGGGGAACGGA 1818
Db      584 AlaHisAspPheIleSerGluLeuArgGluGlyTyrAsnThrIleValGlyGluGlnGly 603
QY      1819 CTCAGGCTTCAGAGAGAAAGCAAAAGAGTATGCAATTCAGAGCCATTTTGAAGAC 1878
Db      604 AlGlyLeuSerGlyGlyGlnArgGlnArgIleAlaIleAlaArgAlaLeuValAsnAsn 623
QY      1879 CCCCAGTCACTACTATATGATGAGCTTCACTTCACTTCACTTCACTTCACTTCACT 1938
Db      624 ProGlyIleLeuIlePheAspGlnAlaThrSerAlaLeuAspTyrGluSerGlnHisVal 643
QY      1939 ATTCTTGTCCTCAGAGAGATGTCGTCACAAACAGAACTTCTATTTCATTCAGACACAGA 1998
Db      644 IleMetArgAsnMetHisIleCysIleCysGlyArgThrValIleIleIleAlaHisArg 663
QY      1999 TTGTCAACAGTGGTTGATGAGATGAAATCATGCTTGGATCGAGGTAAAGTAAAGCCGA 2058
Db      664 LeuSerThrValIleAsnAlaAspArgIleIleValMetGluIleGlyIleValGlu 683
QY      2059 CGTGTATCCCAAGCATGTTGCTTGTCTTCAACCTCATATGATCTTATTCAGAAATGGCAT 2118
Db      684 GlnGlyLeuHisIleGlySerGluLeuLeuSerGluProGluSerLeuTyrSerTyrLeuTyrGln 703
QY      2119 ACAAGAGC 2127
Db      704 LeuGlnSer 706

RESULT 10
AB10_MOUSE STANDARD; PRT; 715 AA.
ID AB10_MOUSE
AC Q9J139; Q9D0C7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family B, member 10, mitochondrial precursor
DE (ATP-binding cassette transporter 10) (ABC transporter 10 protein)
DE (ABC-mitochondrial erythroid protein) (ABC-me protein).
GN ABCB10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Erythrocyte;
RX PubMed=10835348;
RA Shirihai O.S., Gregory T., Yu C., Orkin S.H., Weiss M.J.;
RT "ABC-me: a novel mitochondrial transporter induced by GATA-1 during
RT erythroid differentiation."
RL EMBO J. 19:2492-2502(2000).
RP
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saio T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saio R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batsalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsumoto Y., Nikaide Y., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

```











D 534 LeuAlaLeuAlaLeuProAbenTPrLeuArgGlnIleGlyValValLeuGlnAspAsn 553  
Q 1666 GTCCCTTCATTAATACATATTATTATTAACAACCTCTTAATAGGAACATCAGTCTTCACT 1725  
D 554 ValLeuLeuLeuLeuArgSerIleArgGlnAsnIleAlaLeuSerAspProGlyMetPromet 573  
Q 1726 GAGAAAGTATGAGAGTGGCAAAATTAGCTGGACTTCAATGATGCAATTTCTGAAATGCCA 1785  
D 574 GluArgValIleTyrAlaIleAlaLeuAlaGlyAlaIleAspPheIleSerGlnLeuArg 593  
Q 1786 CATGATATATGACACCCAGTGGGGAAGAGACTCAAGCTTTTATGAGAGGAAAGCA 1845  
D 594 GluGlyTyrAsnThrIleValGlyGlnGlnIleAlaGlyLeuSerGlyGlyGlnArgGln 613  
Q 1846 AGATAGCAATTCAGAGAGCCATTTTGAAGACCCCCAGTCACTACTATGATGAAGCT 1905  
D 614 ArgIleAlaIleAlaArgAlaLeuValAlaAsnProIleAlaIlePheAspGlnAla 633  
Q 1906 ACTTCATCGTTAGTTCGATTACTGAAAGACATATCTTGGTCCATGAGAGATGGTC 1965  
D 634 ThrSerAlaLeuAspTyrGlnSerGlnIleIleIleMetGlnAsnMetGlnIleCys 653  
Q 1966 AAACAGCAGACTTCATTTTCATTCAGACAGATTCACAGTGTGATGACAGATGA 2025  
D 654 GlnGlyArgThrValIleLeuIleAlaIleArgLeuSerThrValIleAsnAlaAspArg 673  
Q 2026 ATCATTTCTTGGATCAGGTTAAGTAGCCAGTGTGATCCCACTGATGTTGCTTCT 2085  
D 674 IleIleValMetGlnGlySerGlnIleValGlnGlnIleGlySerIleGlnLeuLeuGln 693  
Q 2086 AACCTCATAGTATCATTCAGAAATGTGGCATACAG 2124  
D 694 AsnSerAsnGlyLeuTyrSerTyrLeuHisGlnLeuGln 706

RESULT 12

Y4GM\_RHISN STANDARD; PRT; 586 AA.  
ID Y4GM\_RHISN STANDARD; PRT; 586 AA.  
AC P55469;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable ABC transporter ATP-binding protein Y4GM.  
GN Y4GM.  
OS Rhizobium sp. (strain NGR234) .  
OC Plasmid sym PNGR234a.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_taxid=394;  
ON [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=97305956; PubMed=9163424;  
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
RA Berret X.,  
RT Molecular basis of symbiosis between Rhizobium and legumes."  
RL Nature 387:394-401(1997) .  
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY, MSBA SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

DR EMBL; AB000075; AAB91687.1; -  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; AAA\_transport.  
DR InterPro; IPR001140; ABC\_transporter.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR Pfam; PF00664; ABC\_membrane; 1.  
DR ProDom; PD000006; ABC\_transport; 1.  
DR SMART; SM00382; AAA; 1.

DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
KW Hypothetical protein; ATP-binding; Transport; Transmembrane; Plasmid.  
FT TRANSMEM 30 50 POTENTIAL.  
FT TRANSMEM 66 86 POTENTIAL.  
FT TRANSMEM 152 172 POTENTIAL.  
FT TRANSMEM 256 276 POTENTIAL.  
FT NP\_BIND 379 386 ATP (POTENTIAL).  
SQ SEQUENCE 586 AA; 64262 MW; 1095DFB82620637 CRC64;

Alignment Scores:

pred. No.: 2,82e-37 Length: 586  
Score: 654.00 Matches: 175  
Percent Similarity: 51.59% Conservative: 100  
Best Local Similarity: 32.83% Mismatches: 226  
Query Match: 15.37% Indels: 32  
DB: 1 Gaps: 13

AF133659 (1-2345) x Y4GM\_RHISN (1-586)

Q 562 ATGCAACAGCAGTTCGATGCTATGATGATCAAGACCTGAGCTGTTTATAC 621  
D 69 ValAlaValThrValAlaIleIlePheAlaValIle-----GlyLeuAlaThrTyr--- 85  
Q 622 GAAGTTGAAATGCAGTATTTGGCAAGTAGCCAGATTCATCCGAGAAATAGCCAA 681  
D 86 ---ValGlnSerIlePheLeuSerIleValGlnAsnIleIleAlaIleThrGlnArg 104  
Q 682 AATGCTTCCTCATCTTCACCAACCTGATGCTGGTTTTCACCTGAGCAACAGCGGA 741  
D 105 ArgLeuPheGlnIleHisValLeuArgGlnGlyLeuSerPheTyrSerIleTyrProSerSer 124  
Q 742 GCTTTATCTAAGGCTTATTTGACAGAGAACAGAGGATATCATGTTTGTCTGAGTCTTG 801  
D 125 GluLeuLeuValArgLeuThrAsnAlaGlnAlaValArgSerValIleGlnLeuVal 144  
Q 802 GATTTAATCTTCTCCATCATGTTTGAAGTAGCTGTCTGATGCTGTTTGTATAC 861  
D 145 ValThrSerPheIleArgAspLeuPheSerLeuMetGlyLeuLeuAlaValMetVal--- 163  
Q 862 AATGCGGTCGCCAGTTGCTTGTGTAACCTT-----GGAACACTGTTGATACATACA 915  
D 164 -----IleGlnIleProLeuLeuSerLeuValSerAlaIleValGlyProGlyAla 180  
Q 916 GCATTCACAGTTTCAGTCACACGCTGAGAACATGATTTAGATTAAGAAATGACAAAGCA 975  
D 181 IleLeuGlyValArgValIleuThrArgIleValArgIleIleMetGlnLeuGlnIleAla 200  
Q 976 GATTAATGATGCAGTATGCTGCTATAGACTCACTGCTGAATTATGAACT----- 1026  
D 201 -----SerIleGlyGlnIleIleGlnSerValIleGlnIleThrSerThrGlyIleArg 217  
Q 1027 ---GTGAAGATTTTAAATATGAAGAATATGAGACACAGATATGATGATTTTGAAG 1083  
D 218 IleValIleAlaPheAlaLeuGlnAspPheMetArgArgIleMetAspIleTyrIleGly 237  
Q 1084 ACGTATGAG-----ACGTCTCATGAAAGTCCCTACTCTGCTGCTATG 1128  
D 238 AspValGlnArgArgAlaAsnSerIleAlaArgLeuGlnAlaIleAspSerProIleMet 257  
Q 1129 CTGAACCTTGGTCAAGTGTATTTTCAAGTGC---GGTTAAACACTAATATGATGCTC 1185  
D 258 GluThrLeuSerGlyPheAlaIleAlaIleGlyValIleAlaLeuSerGlyValLeuValLeu 277  
Q 1186 GCCAGTCAGGAATTTGTGGCAGGTACCTTACTGTTGGAGATCTAGTAATGTGAATGCA 1245  
D 278 ---GlnGlnGlyAsnThrProGlyGlnLeu-----MetSerPheIleThrAla 292  
Q 1246 CTGCTTTTTCAGCTTCATTAACCTGGAACCTTCTGGAACCTGATATATGAGAGACTAGA 1305  
D 293 LeuLeuLeuAlaTyrGln---ProAlaIleArgLeuAlaArgMetArgIleSerLeuGln 311  
Q 1306 CAAGCATCATAGATATGAACACCTGTTTACTCTTACTCAAGGTAGACACCCAAATTA 1365

```

Db      312 SerAlaLeuValGlyValArgMetCysTyrGlnLeuAlaAspHisProIleGluLeuThr 331
Qy      1366 GACAAAGATGAGTATCCCTTCAGATCACACGACGACGATCCGTCCTTGAT 1425
Db      332 GlnValAsnSerAlaIleProLeuProGluGlyPro-----GlyGlnIleArgPheLys 349
Qy      1426 AATGTCATTTTGAATATACATTGAGCCGCAAGAAATCTTATGTAATTCCTTTGAAAGTC 1485
Db      350 AspValAsnPheSerTyrLysAsnGlyGlnArgLeuPheGlnAsnLeuAsnValThrPhe 369
Qy      1486 CTGCGAAGAAAGAAAGTGGCCATTGTAGAGTACTGGCTCAGGAAAGCAATAGTACTG 1545
Db      370 ProAlaGlyLysThrThrAlaLeuValGlyProSerGlyAlaGlyLysSerSerIleIle 389
Qy      1546 AGGCAATATTTCGCTTATGAGCTCAAAAGGTAGATATTATCTGCTGGTCAAAAT 1605
Db      390 AsnLeuIleMetArgLeuTyrAspProAspValGlySerValThrValAspGlyHisAsp 409
Qy      1606 ATACAAGATGTAGAGCTGAGAAAGCTTCGAGGCGAGTGGAGTGTACTTCAGAGATGCT 1665
Db      410 LeuLysAspValThrPheArgSerLeuThrArgAspArgIleGlyPheValGlyGlnAspThr 429
Qy      1666 GTCTCTTCATATATATCTATTATTAACAACCTCTTATATGGAACATAGTGCCTTCACT 1725
Db      430 PheLeuPheSerGlyThrIleLysTyrAsnIleSerLeuGlyArgGluGlyValAspAsp 449
Qy      1726 GAGGAAGGTATGAGTGGCAAAATTAAGTCTGAGCTTCATGCAATCTTCGATGCGCA 1785
Db      450 GlnGluIleIleGlnAlaAlaLysThrAlaAsnAlaHisAspPheIleMetLysMetPro 469
Qy      1786 CATGATATGACACCCCAAGTAGAGGAGCAGAGATCAAGCTTCAGAGAGAAAGCAA 1845
Db      470 HisGlyTyrAspThrGlnValGlyGlnAsnGlyIleLysLeuSerGlyGlyGlnLysGln 489
Qy      1846 AGAGTAGAATGCAAGAGCCATTGTGAAGAGCCGCCCATGATCTATGATGAAAGCT 1905
Db      490 ArgIleThrIleAlaArgAlaMetLeuThrArgAsnAlaGlnIleLeuIlePheAspGluAla 509
Qy      1906 ACTTCATGTTAGATTCGATTAAGAGATCTATTTCTTGGTCGATGAAAGATGTGGTC 1965
Db      510 ThrSerAlaLeuAspSerGlnSerGlnIleArgGlnAlaLeuAlaArgLeuThr 529
Qy      1966 AAACACAAACTCTTATTTTTCATTCACACAGATTGTCAACAGTGGTTGATGACATGAA 2025
Db      530 ArgLysAspGlnThrThrIleMetCysLeuAlaHisArgLeuSerThrValThrAlaAlaAspAsn 549
Qy      2026 ATCATGTTCTTGATCAGAGGTAGAGTACCGCAACGTGTGATCCACCAAGTGTGGTGGCT 2085
Db      550 IleValIleMetGlnGlyGlyGlnValAlaGlnGlnGlyPro---GlnGlyArgLeuLeu 568
Qy      2086 AACCTCATATGATCTATTCAGAAATGTGGCATACAG 2124
Db      569 SerGlnAspGlyValTyrArgArgLeuTyrGlnLeuGln 581

```

```

RT      actinomycetemcomitans."/;
RL      Nucleic Acids Res. 18:5291-5291 (1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=2072446; PubMed=1961107;
RA      Lally E.T., Golub E.E., Kleba I.R., Taichman N.S., Decker S.,
RA      Berthold P., Gibson C.W., Demuth D.R., Rosenbloom J.;
RT      "Structure and function of the B and D genes of the Actinobacillus
RT      actinomycetemcomitans leukotoxin complex.";
RL      Microb. Pathog. 11:111-121 (1991).
CC      -1- FUNCTION: INVOLVED IN THE EXPORT OF LEUKOTOXIN (HEMOLYSIN).
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein. HLXB SUBFAMILY.
CC      -1- SIMILARITY: BEHONGS TO THE ABC TRANSPORTER FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL, X53955; CA37906.1; -.
DR      PIR, S12601; S12601.
DR      PIR, A61378; A61378.
DR      MEROPS, C39. UNM; -.
DR      InterPro: IPR003593; AAA_ATPase.
DR      InterPro: IPR003439; ABC_transport.
DR      InterPro: IPR001140; ABCtransportTM.
DR      InterPro: IPR005074; Peptidase_C39.
DR      Pfam: PF00005; ABC_tran; 1.
DR      Pfam: PF00664; ABC_membrane; 1.
DR      Pfam: PF03412; Peptidase_C39; 1.
DR      ProDom: PD000006; ABC_transport; 1.
DR      SMART, SM00382; AAA; 1.
DR      PROSITE, PS00211; ABC_TRANSPORTER; 1.
DR      HEMOLYSIS; Transport; ATP-binding; Transmembrane.
DR      NP_BIND, 502, 509, ATP (BY SIMILARITY).
SQ      SEQUENCE 707 AA; 79578 MW; CA564EACF46DF4FB CRC64;

Alignment Scores:
Pred. No.: 4.27e-37 Length: 707
Score: 651.50 Matches: 173
Percent Similarity: 52.13% Conservative: 121
Best Local Similarity: 30.67% Mismatches: 233
Query Match: 15.31% Indels: 37
DB: 1 Gaps: 13

AF133659 (1-2345) x HLXB_ACTAC (1-707)
Qy      463 ATTAGGTTCCCTCATGTTTAATATGCTGTAGACAGAGCTCAACAGATGTCGGA--- 519
Db      169 leuIleThrProLeuPhePheGlnValValMetCysPylsValLeuValHisArgGlyPhe 188
Qy      520 AACATGCTGAACCTGATGATGACCAAAATACAGTTGCAACATGAGGACAGTTCG 579
Db      189 SerThrLeuAsn-----ValIleThrValAlaLeuAlaIleVal 201
Qy      580 ATTAGGCTATGCTGTATCAAGAGCTGAGAGCTGCTTTTAAACGAAGTTCCGAATGCGATA 639
Db      202 ValLeuPheGlnIleIleLeuGlyGly-----LeuArgThrTyrVal 215
Qy      640 TTGGCAAGGTAGGCCCAATTCATCCGAAGATACCAAAATGCTTTCTCCATCTT 699
Db      216 PheAlaHisSerThrSerArgIleAspValGlnLeuGlyAlaArgLeuPheArgHisLeu 735
Qy      700 CACAACCTGATGCTGGTGTTCACCTGACAGACAGACGAGGAGCTTATCTTAAGGCTATT 759
Db      236 leuAlaLeuProIleSerTyrPheGlnAlaArgArgValGlyAspThrValAlaArgVal 255
Qy      760 GACAGAGCAACAGAGGGGTATCAGTTTGTCTGAGTCTGTGATTTAATCTTCTCC 819
Db      256 ArgGlnLeuAspGlnIleArgAsnPheLeuThrGlyGlnAlaLeuThrSerIleLeuAsp 275

```

QY 820 ATCAATGTTGAAGTGGCTTGTGAGTGGTGTGTTGATTAACAATGGCGTCCCACTTT 879  
Db 276 LeuLeuPheSerPheIlePhePheAlaValMetPheTyr-----SerProLysLeu 293  
QY 880 GCTTTGGTAACCTTGGAAACACTTGTAACAACAGATTCACAGTTGACGTACA--- 936  
Db 294 ThrLeuValValLeuGlySerLeuProCysTyrValIleTyrSerValPheIleSerPro 313  
QY 937 -----CGGTGGAACACTAGATTAGATAGAAATAGAAACAAGACAGATTAATGTCAGGT 990  
Db 314 IleLeuArgAlaGlyLeuAspLysPheAlaArgAsn---AlaAspAsnGln----- 330  
QY 991 AATGCTGCTATGACATCAGCTGCTGAATTAAGAACTGGAAGTATTTTAATATGAAAGA 1050  
Db 331 SerPheLeuValGlnSerValThrAlaIleAsnThrIleLysValMetAlaIleSerPro 350  
QY 1051 TATGACACACAGATATGATGAGATTTTGAAGACGTAATGAGACTGCTTCAATGAAAAGT 1110  
Db 351 GlnMetThrAsnIleTyrAspLysGlnLeuAlaSerTyrValAlaValSerPheLysVal 370  
QY 1111 ACTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1170  
Db 371 Thr-----ValLeuAlaThrIleGlyGlnGlnGlyIleGlnLeuIle---GlnLys 386  
QY 1171 GCTATTAATGCTG-----CTGCGCAGTCAGGGAATTTGCGAGGTACCTTACT 1218  
Db 387 AlaValMetValIleAsnLeuTyrLeuGlyAlaHisLeuValIleSerGlyAspLeuSer 406  
QY 1219 GTTGGAGATCTAGTAATGTCGTAATGACTGCTTTTCACTTCACTTCACTTCACTT 1278  
Db 407 IleGlyGlnLeuIleAlaPheAsnMetLeuAlaGlyGlnIleIleSerProValIleArg 426  
QY 1279 CTGGAACCTGATATAGAGACTAGACACAGACCTCAGTATGATGACACCTGTTTACT 1338  
Db 427 LeuAlaGlnIleTyrGlnAspPheGlnGlnValGlyIleSerValThrArgLeuGlyAsp 446  
QY 1339 CTACTCAAGTAGACACCCCAATTAAAGACAAGTAGTCGATCTCC---CTTCAGATC 1395  
Db 447 ValLeuAsnSerProThrGlnAsnAsnThrAlaSerValSerLeuProGlnIleGln--- 465  
QY 1396 ACACCAACAGACACTACCGGCTTGTGATATGTCATTTGAAATAC--ATTGAGGCG 1452  
Db 466 -----GlyGlnIleSerPheArgAsnIleLysPheArgTyrLysProAspSer 481  
QY 1453 CAGAAAGCTTATGGAATATCTTTGAAGTCCCTGACGAAAGAAAGTGGCCATTTGA 1512  
Db 482 PrometIleLeuAsnAsnIleAsnLeuAspLleSerGlnGlyValIleGlyIleVal 501  
QY 1513 GAGAGTAGTGGTCAAGGAAACACAAATAGTACGCTATTTTCCCTTATATGAGCCT 1572  
Db 502 GlyArgSerGlySerGlyLysSerThrLeuThrLysLeuIleGlnArgPheTyrIlePro 521  
QY 1573 CAAAAGGTACATTATCTTGTCTGTCAAATAATACAAAGTGTGACCTTGAAGAACTTT 1632  
Db 522 GlnGlnGlnGlnValLeuIleAspGlyHisAspLeuAlaLeuAlaAspProAsnTyrLeu 541  
QY 1633 CGAGGCGCAGTGGGAGGTGATACCTCAGATCTCTCTTCCCTATATCTATTTATAC 1692  
Db 542 ArgArgGlnValGlyValIleValLeuGlnAspAsnValLeuLeuAsnArgSerIleArgGln 561  
QY 1693 AACCTCTTATATGGAACATCAGTCTTACCTGAGAGGTGATGAGGCGGCAAAATTA 1752  
Db 562 AsnIleAlaLeuThrAsnProGlyMetProMetGlnLysValIleAlaAlaAlaLysLeu 581  
QY 1753 GCTGGACTTATGATGCAATTTCTTGAATGCCACATGATATGACACCAAGTAGGGGAA 1812  
Db 582 AlaGlyAlaHisAspPheIleSerGlnLeuArgGlnGlyTyrAsnThrValIleGlyGln 601  
QY 1813 CGAGGACTCAGCTTATGAGAGGAAAGAAAGACTATGCAATTTGGAAGACCATTTTG 1872  
Db 602 GlnGlyAlaGlyLeuSerGlyGlnArgGlnArgIleAlaIleAlaArgAlaLeuVal 621

QY 1873 AAGACCCCCCACTATCTATGATGAAGCTACTCATGCTTGAATTCAGTACTGAA 1932  
Db 622 AsnAsnProArgIleLeuIlePheAspGlnAlaThrSerAlaLeuAspTyrGlnSerGln 641  
QY 1933 GAGACTATCTTGGTGGCAGGAAGATGTGCTCAACACAGACTTGTATTTTATGCA 1992  
Db 642 AsnIleLeuMetHisAsnMetHisLysIleCysGlnAsnArgThrValLeuIleAla 661  
QY 1993 CACAGATTGTCACACAGTGTGATGACAGATGAATCATGCTTGTGATCAGGTAAAGTA 2052  
Db 662 HisArgLeuSerThrValLysAsnAlaAspArgIleIleValMetAspLysGlyIle 681  
QY 2053 GCCGAACGTGTACCCACCAATGCTTGTGCTTAACCTCATGATATCATGAAATG 2112  
Db 682 IleGlnGlnGlyLysHisGlnGlnLeuLeuLysAspGlnLysGlyLeuTyrSerTyrLeu 701  
QY 2113 TGGCATACACAG 2124  
Db 702 HisGlnLeuGln 705  
RESULT 14  
MSBA\_ECOLI STANDARD, PRT; 582 AA.  
ID MSBA\_ECOLI  
AC P27239;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable transport ATP-binding protein mbda.  
GN MSBA OR B0914 OR Z1260 OR ECS0997.  
OS Escherichia coli, and  
OC Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxId=562, 83334;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=93172962; PubMed=8094880;  
RA Karow M.L., Georgopoulos C.P.;  
RT "The essential Escherichia coli mbda gene, a multicoopy suppressor of  
RT null mutations in the hcrB gene, is related to the universally  
RT conserved family of ATP-dependent translocators.";  
RL Mol. Microbiol. 7:69-79(1993).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12; MG1655;  
RX MEDLINE=9742617; PubMed=9278503;  
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshina T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kikagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
RN [4]  
RN SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grobbeck E.J., Davis N.W., Lim A., Dimailanta E.T., Potamoculis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIND 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsuo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Kida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,  
 RA Kitaura S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 RL [6]  
 RP CHARACTERIZATION.  
 RA MEDLINE=98241619; PubMed=9575204;  
 RX Zhou Z., White K.A., Polissi A., Georgopoulos C., Raetz C.R.;  
 RT "Function of *Escherichia coli* MsbA, an essential ABC family  
 transporter, in lipid A and phospholipid biosynthesis.";  
 RL J. Biol. Chem. 273:12466-12475(1998).  
 RL [7]  
 RP X-RAY CRYSTALLOGRAPHY (4.5 ANGSTROMS).  
 RX Chang G., Roth C.B.;  
 RT "Structure of MsbA from *E. coli*: a homolog of the multidrug resistance  
 ATP binding cassette (ABC) transporters.";  
 RL Science 293:1793-1800(2001).  
 CC -1- FUNCTION: INVOLVED IN THE SECRETION OF LIPID A AND POSSIBLY ALSO  
 CC GLYCEROPHOSPHOLIPID TRANSPORT.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MSBA SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL; 211796; CAA7839.1; -.  
 DR EMBL; AE000193; AAC74000.1; -.  
 DR EMBL; D90729; BA35658.1; -.  
 DR EMBL; D90730; BA35660.1; -.  
 DR EMBL; AE005281; AAG55399.1; -.  
 DR EMBL; AP002553; BAB34420.1; -.  
 DR PIR; S21588; S21588.  
 DR PIR; S27998; S27998.  
 DR PDB; 1J50; 12-SEP-01.  
 DR EcGene; EG10613; mbaA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transport.  
 DR InterPro; IPR001140; ABCtransport.  
 DR Pfam; PF00005; ABC\_tran.1.  
 DR Pfam; PF00664; ABC\_membrane.1.  
 DR SMART; SM00382; AAA.1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER.1.  
 KW ATP-binding; Transport; Inner membrane; Transmembrane; 3D-structure;  
 KW Complete proteome.  
 FT DOMAIN 1 21 CYTOPLASMIC.  
 FT TRANSMEM 22 52  
 FT DOMAIN 53 64 EXTRACELLULAR.  
 FT TRANSMEM 65 96  
 FT DOMAIN 97 139 CYTOPLASMIC, INTRACELLULAR DOMAIN 1.  
 FT TRANSMEM 140 164  
 FT DOMAIN 165 167 EXTRACELLULAR.  
 FT TRANSMEM 168 192  
 FT DOMAIN 193 252 CYTOPLASMIC, INTRACELLULAR DOMAIN 2.  
 FT TRANSMEM 253 272  
 FT DOMAIN 273 280 EXTRACELLULAR.

FT TRANSMEM 281 301  
 FT DOMAIN 302 582 CYTOPLASMIC, INTRACELLULAR DOMAIN 3.  
 FT NP\_BIND 376 383 ATP.  
 SQ SEQUENCE 582 AA; 64460 MW; 3D52A3482174A7FE CRC64;  
 Alignment Scores:  
 Pred. No.: 4.51e-37 Length: 582  
 Score: 651.00 Matches: 190  
 Percent Similarity: 49.19% Conservative: 115  
 Best Local Similarity: 30.65% Mismatches: 259  
 Query Match: 15.30% Indels: 56  
 DB: 1 Gaps: 17  
 AF133659 (1-2345) x MSBA\_ECOLI (1-582)  
 QY 310 CTCACACAGACCCAAAGAGGTTAAAGATGTGATCTGCAAAATCATTAAGCA 369  
 DB 1 Methisambsp-----LysAspLeuSerThrTrpGlnThrPheArgArg 15  
 QY 370 ATGCTTCTTATGTGTGGCCCAAGACAGCCAGAT-----CTACGAGTGA 417  
 DB 16 Leu-----TTProThrIleAlaProPheLysAlaGlyLeuValAlaGly 31  
 QY 418 GTTGCCATTGCTGGGATTTTGGTGCGCAAGCCATGAATATTGCTTCCTTC 477  
 DB 32 ValAlaLeuIleLeuAsnAlaLaserAspThrPheMetLeuSerLeuLysProLeu 51  
 QY 478 ATGTTTAAATATGTGTGATGACAGCTCAACAGATGCGGAAACATGCTGAAGT 537  
 DB 52 Leu-----AspAsp-----GlyPheGly 57  
 QY 538 GATGACCAATATGATGATGATGACCATGCAACAGATGCTGATGCTATGATCA 597  
 DB 58 LysThrAspArgSerValIleValTrpMetProLeuValIleGlyLeuMetIleu 77  
 QY 598 AGAGCTGAGCTGCTTTTAAACGAA---GTGGAATGACGATTTTGGACGTACC 654  
 DB 78 ArgGlyIleThrSerIleValSerSerIleValSerIleValSerIleValSerIleVal 97  
 QY 655 CAGATTCATATCCAGAGATGCAAAATATGCTTTCATCTTCAACACCTGATCTG 714  
 DB 98 MetThrMetAlaArgValGlu-----PheGlyIleMetMetGlyMetProVal 113  
 QY 715 GGTTCACCTGACAGACAGACGCGAGCTTATCTAAGCTATGACAGAAACAGG 774  
 DB 114 SerPhePheAspLysGlnSerThrGlyThrLeu-----LeuSerArgIleThrTyr 130  
 QY 775 GGTATCAGTTTGTCTGAGTCTTGGTATTTATCTTCTCCATCATGTTGAAGTG 834  
 DB 131 AspSerGlnValAlaLaserSerSerGlyAlaLeuIleThrValAlaArgGly 150  
 QY 835 ATGCTGTCACTGTG-----GTTTGATTAACAAGCGGCGCCACTTGCT 882  
 DB 151 AlaSerIleIleGlyLeuPheIleMetMetPheTyrTrpSerTrpGlnLeuSerIleIle 170  
 QY 883 TTGTGAACCTTGAACACTTGTATACATACACAGATTCACAGTGCACAGCGTGG 942  
 DB 171 LeuIleValLeuAlaProIle-----ValSerIleAlaIleArgVal 185  
 QY 943 AGAATGATTTAGATTAAGAAATGAACAGACAGATTAATGACAGT-----AAT 993  
 DB 186 SerLysArgPheArgSerIleSerLysAsnMetGlnAsnThrMetGlyGlnValThr 205  
 QY 994 GCTGCTATAGACTCACTGCTGAATTAATGAAGAGTGAAGTATTAATGAAGAATAT 1053  
 DB 206 SerAlaGlnGlnMetLeuLysGlyIleLysGlnValLeuIlePheGlyGlnGlnVal 225  
 QY 1054 GAACACACAGATATGATGATTTTGAAGACGTATGACAGTCTTCAATGAAGAAC 1113  
 DB 226 GluThrLysArgPheAspLysValSerAsnArgMetArgLeuGlnGlyMetLysMetVal 245  
 QY 1114 TCTACTGCTATGCTGAAGCTTTGTCAAAGTCTATTTTCAGTGTGCTTTAAACAGCT 1173

|            |   |  |   |              |
|------------|---|--|---|--------------|
| Db         | 246   | Seal   | AlaSerSerIleSerAspProIleIleGlnLeuIleLeuIleSerLeuLeuAlaLeuAlaPhe | 265          |
| QY         | 1174  | ATAATGTCCTCGCCAGTCAGGGAAATTGTGCAAGTACCTTACTGTGTGAGACTCACTA         | 1233  |              |
| Db         | 266   | ValLeuIrrAlaIleAspPheProSerValMetAspSerLeuThrIleGlyThrIleThr       | 285   |              |
| QY         | 1234  | ATGGGAAATGACACTGCTTTTACGCTTCAATTACCTGAACTTTGTGGAACTGTATAT          | 1293  |              |
| Db         | 286   | ValValPheSerSerMetIleAlaLeuMetArgProLeuSerLeuThrAsnValAsn          | 305   |              |
| QY         | 1294  | AGAGACATTAACACAGACATCATGATATATGAACCTGTGTTACTCTTACCAAGTAGAC         | 1353  |              |
| Db         | 306   | AlaGlnPheGlnArgGlyMetAlaIaIaCysGlnThrLeuPheThrIleLeuAspSerGln      | 325   |              |
| QY         | 1354  | ACCCAAATTAAGA CAAAGTAGATGGACTCTCCCTCAATGTCACACACAGACAGTACC         | 1413  |              |
| Db         | 326   | --GlnGlnIlyAspGln-----GlyArgValIleGlnArgIleThrGlyAsp               | 341   |              |
| QY         | 1414  | GTGGCTTGTATATGTGCATTTTGAATATACATTAGAGGCCAAGAAAGTC-----CTTACT       | 1467  |              |
| Db         | 342   | ValGlnPheArgAsnValThrPheThrTyr--ProGlyArgAspValProAlaLeuArg        | 360   |              |
| QY         | 1468  | GGAAATTCCTTGAAGTCCCGCAGGAAAGAAAGTGCCCATTTGTAGAGTAGTGTGGATCA        | 1522  |              |
| Db         | 361   | AsnIleAsnLeuIyrIlePheAlaGlyIlySerThrValAlaLeuValGlyArgSerGlySer    | 380   |              |
| QY         | 1528  | GGGAAAGACAAATATGTAGAGCTATATATTCGCTTATAGAGCTCAAAAGGTAGCATT          | 1587  |              |
| Db         | 381   | GlyIysSerThrIleAlaSerLeuIleThrArgPheTyrAspIleAspIleGlnIle          | 400   |              |
| QY         | 1588  | TATCTTGTGCTCAAAATATACAAAGATGTAGGCTGGAAAGCTTTGGAGGGCAAGTGGGA        | 1647  |              |
| Db         | 401   | LeuMetAspGlyHisAspLeuArgGlnIlyThrLeuAlaSerLeuArgAsnGlnValAla       | 420   |              |
| QY         | 1648  | GTGGACCTCAGAGTGCATGCTCTTCACAAATACATATTTATTAACACCTCTTAATGGA         | 1707  |              |
| Db         | 421   | LeuValSerGlnAsnValIleHisIlePheAsnAspThrValAlaAsnAsnIleAlaTyrAla    | 440   |              |
| QY         | 1708  | AACATCACTGCT--TCACCTGAGGAAGTATATGCAGTGGCAAAATTAAGCTGACTTCAT        | 1764  |              |
| Db         | 441   | ArgThrGlnGlnIlyrSerArgGlnGlnIleGlnGlnAlaIaArgMetAlaTyrAlaMet       | 460   |              |
| QY         | 1765  | GATGCAATTCCTTGAATGCCACATGATATATGACACCCAAAGTAGGGAGCAGAGACTCAAG      | 1824  |              |
| Db         | 461   | AspPheIleAsnIlySmetAspAsnGlyLeuAspThrValIleGlyAsnGlyValLeu         | 480   |              |
| QY         | 1825  | CTTTCAGAGAGGAAAGCAAGATGTAGCAATTCAGAGGCCATTTTGAAGACCCCCCA           | 1884  |              |
| Db         | 481   | LeuSerGlyGlyGlnArgGlnArgIleAlaIleAlaArgAlaLeuLeuArgAspSerPro       | 500   |              |
| QY         | 1885  | GTCATACTCTATGTAGGAAGTACTCTTCACTCTTGATTCATTAAGTCAAGACTTATCTT        | 1944  |              |
| Db         | 501   | IleLeuIleLeuAspGlnIleAlaThrSerAlaLeuAspThrGlnSerGlnArgAlaIleGln    | 520   |              |
| QY         | 1945  | GTCGCCATGAAGAGTGTGTCAAAACAGAACTTCATATTTTCATTGCACACAGATTGTCA        | 2004  |              |
| Db         | 521   | AlaIleLeuAspGlnIleGlnIlyAsnArgThrSerLeuValIleAlaHisArgLeuSer       | 540   |              |
| QY         | 2005  | ACAGGTGTATATGTAGATGAATATCATTTGTCTTGGAACAGGATGAAGTACCGCAAGTGTGT     | 2064  |              |
| Db         | 541   | ThrIleGlnIlyAlaAspGlnIleValIleValIleValGlnAspGlyValIleValGlnArgGly | 560   |              |
| QY         | 2065  | ACCCACCATGGTTGTGCTGTGCAACCCATAGATATCTATTCAGAAATGTGGCATCAAG         | 2124  |              |
| Db         | 561   | ThrIleAsnAspLeuLeuGlnHis--ArgGlyValTyrAlaGlnLeuHisIlySmetGln       | 579   |              |
| RESULT 15  |   |  |   |              |
| HLYP_PASSP | ID  | HLYP_PASSP   | STANDARD;   | PRT; 708 AA. |
| AC         | P51122;                                       |  |   |              |
| DT         | 01-OCT-1996 (Rel. 34, Created)                |  |   |              |
| DT         | 01-OCT-1996 (Rel. 34, Last sequence update)   |  |   |              |
| DT         | 16-OCT-2001 (Rel. 40, Last annotation update) |  |   |              |



QY 235 -----GATGCTGAAAGGCTCTCCAGTATGCGCACTGATAGAA 273  
DB LeuThrTyrAsnLeuGluGlnAspAlaProlysIleLeuSer-----GlnGluGlu 114  
QY 274 AAGAGACATGTTGGCATGCTGTCACAGAGAGAGACTCCACACAGACCCAAAGAGGG 333  
DB PheGluSerCysTyrGlnGlyLys----- 122  
QY 334 TTTAAAGATGTGATACCTCGGAAATCATTAAGCAATGCTTTCT- -TATGCG 384  
DB 123 IleIleLeuValThrSerArgAlaSerIleValGlyGlnLeuAlaLysPheAspPheThr 142  
QY 385 TGG-----CCCAAGACAGCGCCAGATCTACAGAGCTAGAGTTGCCATTTG 429  
DB 143 ThrPheIleProAlaValIleLysTyrArgLysIlePheLeuGluThrLeuLeuValSer 162  
QY 430 CTGGGATTTTGGTGGTGGCAAAAGCCATGAATATGTGTTCCCTTCATGTTAAATAT 489  
DB 163 Ile--PheLeu-----GlnIlePheAlaLeuIleThrProLeuPheGlnVal 178  
QY 490 GCTGAGACAGCCCTCAACCAATGTCGGA--AACATGCTGAACCTGAGTATGACCA 546  
DB 179 ValMetAspLysValLeuValHisArgLysPheSerThrLeuAsn----- 193  
QY 547 AATACAGTTGCAACCATGCAACAGCAGTTCGATTTGCTATGCTATCAAGACTGGA 606  
DB 194 -----IleIleThrValAlaLeuAlaIleValIleIlePheGluIleValLeuSerGly 211  
QY 607 GCTGCTTTTAAAGAAAGTTCGAAATGCAATGATTTGGCAGAGTACCAGAAATTCATC 666  
DB 212 -----LeuArgThrTyrIlePheAlaHisSerThrSerArgIleAsp 225  
QY 667 CGAAGATATGCAAAATATGCTTTCCATCTTCACAACTGAGATCTGGTTTACCTG 726  
DB 226 ValGluLeuGlyAlaArgLeuPheArgHisLeuLeuAlaLeuProIleSerTyrPheGlu 245  
QY 727 AGCAGACAGACGAGAGCTTATCTTAAGGCTATGACAGAAACAAAGGATTCAGTTT 786  
DB 246 AsnAspArgValGlyAspThrValAlaArgValArgLysLeuAspGlnIleArgAsnPhe 265  
QY 787 GTCCTGAGTGTGCTGATTAATTAATCTTCCATCATGTTTGAAGTATGCTTCACT 846  
DB 266 LeuThrGlyGlnAlaLeuThrSerValLeuAspLeuPheSerPheIlePhePheAla 285  
QY 847 GGTGTTTGTATTAACAATGCGGTGCCAGTTCGTTGGTAACTCCCTGGAACACTGGT 906  
DB 286 ValMetTyrTyr-----SerProLysLeuThrLeuValIleLeuGlySerLeuPro 303  
QY 907 ACATACACAGCATTCACAGTTGCAGTTCACACGCTGAGAACTAGATTGAATAGAA-- 963  
DB 304 CysTyrIleLeuThrSerIlePheIleSerProIleLeuArgArg--ArgLeuAspAsp 322  
QY 964 -----ATGAACAAAGCAGATATGATGAGTAAATGCTGCTATAGACTCACTGCTGAAT 1017  
DB 323 LysPheAlaArgGlyAlaAspAsnGlnAla-----PheLeuValGlySerValThrAla 340  
QY 1018 TATGAACCTGCAAGTATTTTAATGAAAGATATGAAGCACAAGATATATGATTT 1077  
DB 341 IleAsnMetIleLysAlaMetAlaValSerProGlnMetThrAspThrTyrAspLysGln 360  
QY 1078 TTGAAGAGGTATGAGACGCTCATTTGAAAGTACCTGACTGCTGCTATGTAACCTT 1137  
DB 361 LeuAlaSerTyrValSerSerSerPheArgValThr-----ValLeuAlaThrIle 377  
QY 1138 GGTCAAGCTGATTTTTCAGTGTGCGATTTACAGCTATATAGTG- -CTCGCC 1188  
DB 378 GlyGlnGlnGlyValGlnLeuIleGlnLysThrValMetValIleAsnLeuThrLeuGly 397  
QY 1189 AGTCAGGAAATTTGGCAGGTACCTTACTGTTGAGATCTAGTAAATGCTAATGACTG 1248  
DB 398 AlaHisLeuValIleSerGlyAspLeuSerIleGlyGlnLeuIleAlaPheAsnMetLeu 417

QY 1249 CTTTTCAGCTTTTCATTACCCCTGAACTTTCTGGAACTGATATGAGACTAGACAA 1308  
DB 418 SerGlyGlnValIleAlaProValIleArgLeuAlaGlnLeuThrGlnAspPheThrGln 437  
QY 1309 GCATCTCATGATATGAACACCTTGTATTACTTACTCAAGTATGACAC--CAATTTAA 1365  
DB 438 ValGlyIleSerValThrArgLeuGlyAspValLeuAsnSerProThrGlnGlnTyrGln 457  
QY 1366 GACAAAGTATGATGCTCTCCCTTCAGATCAACACAGACAGCTACCGTCCCTTGAT 1425  
DB 458 GlyLysLeuSerLeuProGlnIleGln-----GlyAspIleAlaPheLys 472  
QY 1426 AATGTGATTTTGAATAC--ATTAGAGCCGAAAGTCTTACTGGAATATCTTTGAA 1482  
DB 473 AsnIleArgPheArgTyrLysProAspAlaProThrIleLeuAsnValAsnLeuGln 492  
QY 1483 GTCCCTCAGAAAGAAAGTGGCCATTTGTAGAGTATGGTGGTGAAGGAAAGCATA 1542  
DB 493 IleLysLysGlyGluValIleGlyIleValGlyArgSerGlySerGlyLysThrLeu 512  
QY 1543 GTGAGGCTATATTTCGCTTCTATAGAGCTTCAAAGGGTATGATTTATCTGTGCTCAA 1602  
DB 513 ThrLysLeuLeuGlnArgPheTyrIleProGlnAsnGlyGlnValIleAspGlyHis 532  
QY 1603 AATATACAGATGTGAGCTGGAAGGCTTCGAGGCGAGTGGAGTGTACTTCAGAT 1662  
DB 533 AspLeuAlaLeuAlaAspProAsnThrLeuArgGlnArgIleGlyValIleGlnAsp 552  
QY 1663 GCTGCTCTTCCATATACATTTATTAACAACCTCTTATATGAAACATCAGGCTTCA 1722  
DB 553 AsnValIleLeuAsnArgSerIleArgGlnAsnIleAlaLeuSerGluProGlyMetSer 572  
QY 1723 CCTGAGGAAGTATGACAGTGGCAAAATTAAGCTGCACTTCATGCAATCTTCGAT 1782  
DB 573 MetGluArgValIleTyrAlaAlaLysLeuAlaGlyAlaHisAspPheIleSerAspVal 592  
QY 1783 CCACATGATATGACACCCAAAGTAGGGGAAAGAGACTCAAGCTTTCAGAGAGAAAG 1842  
DB 593 ArgGlnGlyTyrAsnThrIleValGlyGlnGlnIleValGlyLeuSerGlyGlnArg 612  
QY 1843 CAAGAGTACGATTTGCAAGAGCATTGTAAGAACCCCCAGTCATATCTATGATGAA 1902  
DB 613 GlnArgIleAlaIleAlaArgAlaLeuValAsnAsnProLysIleLeuIlePheAspGlu 632  
QY 1903 GCTACTTCATGCTAGATTCATTAAGAGACTATTTCTGTCGCATGAGAGATGT 1962  
DB 633 AlaThrSerAlaLeuAspTyrGluSerGlnHisIleIleMetGlnAsnMetGlnLysIle 652  
QY 1963 GTCAACACAGAACTTCTATTTTCATTTGACACAGATTTGCAACAGTGTGATGCAT 2022  
DB 653 CysGlnGlyArgThrValIleLeuIleAlaHisArgLeuSerThrValLysAsnAlaAsp 672  
QY 2023 GAAATCATTTGCTTGATCAGGGTAAAGGTAGCCAGAGTGTATCCCACTGTTGCTT 2082  
DB 673 ArgIleIleValMetGluLysGlyGlnIleValGlnGlnGlyLysHisAsnGluLeu 692  
QY 2083 GCTAACCTCATGATCATTTTCAAAATGCGATACAG 2124  
DB 693 GlnAsnAsnGlyLeuTyrSerTyrLeuHisGlnLeuGln 706

Search completed: February 15, 2003, 05:54:54  
Job time : 48 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2003, 05:51:35 ; Search time 46 Seconds  
(without alignments)  
9801.531 Million cell updates/sec

Title: AF133659  
Perfect score: 4254  
Sequence: 1 ATGGCGCTGCTCGCATGCA.....TATTAAAAAATCATACATT 2345

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p model -DB=x1h  
-Q=/cgr2.1/USFTO\_spool/AF133659/runat\_10022003\_155441\_29535/app\_query.fasta.1.2503  
-DB=PIR\_73 -QMT=fastran -SUFFIX=xpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AF133659 @CGN 1.1 60 @runat\_10022003\_155441\_29535 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEOUTPUT -NEG\_SCORES=0 -WAIT -LONGLOG -DBY TIMEOUT=120  
-MARN TIMEOUT=30 -THREADS=1 -KGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1          | 1657.5 | 39.0        | 677    | 2     | T10657      |
| 2          | 1620.5 | 38.1        | 680    | 2     | T10656      |
| 3          | 1538.5 | 36.2        | 693    | 2     | T39154      |
| 4          | 1516.5 | 35.6        | 690    | 2     | S54211      |
| 5          | 1454.5 | 34.2        | 631    | 2     | AH0315      |
| 6          | 1446.5 | 34.0        | 609    | 2     | G71731      |
| 7          | 1400   | 32.9        | 628    | 2     | AF3438      |
| 8          | 1375   | 32.3        | 629    | 2     | AG2707      |
| 9          | 1375   | 32.3        | 629    | 2     | H97489      |
| 10         | 1247   | 29.3        | 836    | 2     | JEO248      |
| 11         | 1042   | 22.8        | 643    | 2     | C87412      |
| 12         | 970.5  | 21.7        | 830    | 2     | S25198      |
| 13         | 923.5  | 21.5        | 1025   | 2     | T18376      |
| 14         | 915    | 21.5        | 772    | 2     | T26313      |

|    |       |      |     |   |        |                     |
|----|-------|------|-----|---|--------|---------------------|
| 15 | 725   | 17.0 | 642 | 2 | F84172 | ABC transport prot  |
| 16 | 709   | 16.7 | 581 | 2 | E83790 | ABC transporter (A  |
| 17 | 708   | 16.6 | 597 | 2 | AD2074 | ATP-binding protei  |
| 18 | 706.5 | 16.6 | 593 | 2 | S75352 | ABC-type transport  |
| 19 | 700   | 16.5 | 832 | 2 | B87673 | ABC transporter, H  |
| 20 | 699.5 | 16.4 | 578 | 2 | A89874 | hypothetical prote  |
| 21 | 687   | 16.1 | 585 | 2 | H69857 | ABC transporter (A  |
| 22 | 678.5 | 15.9 | 602 | 2 | A75590 | ABC transporter, A  |
| 23 | 675.5 | 15.9 | 594 | 2 | S72638 | hypothetical ABC e  |
| 24 | 674.5 | 15.9 | 604 | 2 | F69802 | ABC transporter (A  |
| 25 | 673.5 | 15.8 | 600 | 2 | AD3138 | hypothetical prote  |
| 26 | 673.5 | 15.8 | 600 | 2 | H98149 | ATP-binding transp  |
| 27 | 673.5 | 15.8 | 863 | 2 | C90482 | ABC transporter, A  |
| 28 | 672.5 | 15.8 | 546 | 2 | AG1886 | ATP-binding protei  |
| 29 | 672.5 | 15.8 | 707 | 1 | LE8CB  | hemolysin secretio  |
| 30 | 671.5 | 15.8 | 611 | 2 | AH3540 | ABC transporter AT  |
| 31 | 671.5 | 15.8 | 707 | 2 | S10057 | ABC-type transport  |
| 32 | 670.5 | 15.8 | 584 | 2 | E83767 | ABC transporter (A  |
| 33 | 666.5 | 15.7 | 599 | 2 | D98267 | hypothetical prote  |
| 34 | 666.5 | 15.7 | 599 | 2 | AD3017 | ATP-binding protei  |
| 35 | 664   | 15.6 | 600 | 2 | AI2454 | probable ABC-trans  |
| 36 | 662   | 15.6 | 595 | 2 | T31077 | ABC-type transport  |
| 37 | 662   | 15.6 | 650 | 2 | T17482 | leukotoxin express  |
| 38 | 658   | 15.5 | 708 | 2 | C30169 | probable transport  |
| 39 | 654   | 15.4 | 582 | 2 | AF0614 | toxoin secretion AB |
| 40 | 653   | 15.4 | 893 | 2 | AH2007 | hypothetical prote  |
| 41 | 652   | 15.3 | 607 | 2 | A95911 | leukotoxin express  |
| 42 | 651.5 | 15.3 | 707 | 2 | A61378 | probable ABC-type   |
| 43 | 651   | 15.3 | 582 | 2 | S27998 | ATP-binding transp  |
| 44 | 651   | 15.3 | 582 | 2 | C85617 | ATP-binding transp  |
| 45 | 651   | 15.3 | 582 | 2 | E90753 | ATP-binding transp  |

ALIGNMENTS

RESULT 1  
T10657  
probable ABC-type transport protein T5F17.80 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jan-2000  
C/Accession: T10657  
R/Bevan, M.; Demard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancr  
submitted to the Protein Sequence Database, June 1999  
A/Reference number: Z16533  
A/Accession: T10657  
A/Molecule type: DNA  
A/Residues: 1-677 <BEV>  
A/Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.80  
A/Experimental source: cultivar Columbia; BAC clone T5F17  
C/Genetics:  
A/Gene: ATSP:T5F17.80  
A/Map position: 4  
A/Intons: 519/3  
C/Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology  
F/452-646/Domain: ATP-binding cassette homology <ABC>

Alignment Scores:

|                        |         |               |         |     |
|------------------------|---------|---------------|---------|-----|
| Pred. No.:             | 3       | 52e-112       | Length: | 677 |
| Score:                 | 1657.50 | Matches:      | 326     |     |
| Percent Similarity:    | 74.30%  | Conservative: | 128     |     |
| Best Local Similarity: | 53.36%  | Mismatches:   | 152     |     |
| Query Match:           | 38.96%  | Indels:       | 5       |     |
| DB:                    | 2       | Gaps:         | 4       |     |

AF133659 (1-2345) x T10657 (1-677)  
QY 316 ACAGACCCAAAGAGGGTTAAAGATGTTGATCTCGGAAATCATMAAGCAATGCTT 375  
Db 68 ThnProasndlnspgdnthrlrlysthrAlaserSerlyslleuauGthrllleSer 87  
QY 376 TCTTATGTGTGGCCCAAGACAGCGCATCTACGAGCTGAGTGGCATTTTGCTGGGA 435  
Db 88 SerTyrlenutrlpmetrllysAspaenProgluLeuArpPheArValIleAlaIleuAla 107

[illegible]

|    |      |  |      |
|----|------|--|------|
| Qy | 1507 | ATTGTAGAGATGTTGGGTACGGAAACACAAATGAGAGCTTATTTTGGCTTCAT          | 1566 |
| Db | 467  | ITLVALYSERISERGLYSERGLYSSERITLLELAAAGMETLLEPHEAPHEPHE          | 486  |
| Qy | 1567 | GAGCTCAAAAGGTAGCATTTATCTTGTGTGTCAAATATATCAAGATGTAGCCTGGA       | 1626 |
| Db | 487  | AspHraspsergilyanValargIleaspglyGlnaspiIelysgluValIthrLeucl    | 506  |
| Qy | 1627 | AGCCTTGGAGGGCAGTGGAGTGGTAACTCAGATGCTGTCTCTTCCATTAATCTT         | 1686 |
| Db | 507  | SerLeuIarGerCysIleGlyValValProGlnasPheThrValLeuPheasnaSPThrIle | 526  |
| Qy | 1687 | TATPACACCTCTTATATGGAAACATCAGTGGCTTCACTGAGGAGGAGTATGCAAGTGA     | 1746 |
| Db | 527  | PheIasnaIleIleIstYrIglYanLeuSerIalatrngIuGluValItyrAspIala     | 546  |
| Qy | 1747 | AAATTAAGCTGGACTTCATGATGCAATCTTTCGAATGCCACATGATATGACCCCAAGTA    | 1806 |
| Db | 547  | ArgYrghIlaValIleHisaspThrIleMetLysPheProasplysSerThrIalaVal    | 566  |
| Qy | 1807 | GGGGAACAGAGACTTCAAGCTTTCAGAGAGAAAGCAAAAGTAGCAATTGCAAGACC       | 1866 |
| Db | 567  | GLYGLuIarGlyLeuMetLeuSerGlyGlyGluLysGlnAspValAlaLeuAlaArgIa    | 586  |
| Qy | 1867 | ATTTTGAGAGACCCCCAGCTACTACTATGATGTAAGGTACTTCATCTTATGATTCATT     | 1926 |
| Db | 587  | PheLeuIysSerProalIleLeuLeuGlysaSpgIuIatIrnAsnAlaLeuaspSerLys   | 606  |
| Qy | 1927 | ACTGAAGAGACTTTCCTGGTGCATGAAGATGTGTGTCAAACACAGAACTTATTTTC       | 1986 |
| Db | 607  | ThrGluIaGlnIleMetLysThrPheaspSerIleuAlaSerAsnArgIhCysIlePhe    | 626  |
| Qy | 1987 | ATTGCAACAGATTGTCACAGAGTGTGTGTCAGATGGAATCATTTGTTCGATCAGAGGT     | 2046 |
| Db | 627  | ITLalHisrgrLeuThrIthrIalameGlnCysaSpGluIleIleValmeGluLysGly    | 646  |
| Qy | 2047 | AAAGTACCCGAAGCTGTATCCACCATGCTTGTCTTGTCTAAACCTCATATGATCTATTTCA  | 2106 |
| Db | 647  | LysValValGluLysGlyIthrHisGlnValIleuLeu--GluLysSerGlyArgTyrAla  | 665  |
| Qy | 2107 | GAATATGCGATACACAGAGACCGGTGCGAG                                 | 2139 |
| Db | 666  | LysLeuItrPthrGlnGlnIasnsrThrLeuGlu                             | 676  |

RESULT 2

T10656  
probable ABC-type transport protein TSPF7\_70 - Arabidopsis thaliana

C|Species: Arabidopsis thaliana (mouse-ear cress)

C|Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jan-2000

C|Accession: T10656

R|Beyan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft, J.

submitted to the Protein Sequence Database, June 1999

A|Reference number: Z16533

A|Accession: T10656

A|Molecule type: DNA

A|Residues: 1-680 <EMBL>

A|Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:TSPF7\_70

A|Experimental source: cultivar Columbia; BAC clone TSPF7

C|Genetics:

A|Gene: ATSP.TSF17.70

A|Map position: 4

A|Introns: 522/3

C|Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology <ABC>  
/455-649/Domain: ATP-binding cassette homology <ABC>

|                        |           |
|------------------------|-----------|
| Alignment Scores:      |           |
| Pred. No.:             | 1,69e-109 |
| Score:                 | 1620.50   |
| Percent Similarity:    | 71.32%    |
| Best Local Similarity: | 50.87%    |
| Query Match:           | 38.09%    |
| DB:                    | 2         |
|                        |           |
| Length:                | 686       |
| Matches:               | 321       |
| Conservative:          | 129       |
| Mismatches:            | 170       |
| Indels:                | 11        |
| Gaps:                  | 5         |





A:Accession: S54211  
 A:Molecule type: DNA  
 A:Residues: 1-690 <KIS>  
 A:Cross-references: EMBL:X81715; NID:G793876; PIDN:CAAS7359.1; PID:G793877  
 R:Conor, R.; Churcher, C.M.  
 submitted to the EMBL Data Library, April 1995  
 A:Reference number: S53969  
 A:Accession: S53971  
 A:Molecule type: DNA  
 A:Residues: 1-690 <CON>  
 A:Cross-references: EMBL:Z49212; NID:G798940; PIDN:CAAB9134.1; PID:G798943; MIPS:YMR301C  
 R:Leighton, J.; Schatz, G.  
 EMO J. 14, 186-195, 1995  
 A:Title: An Aac transporter in the mitochondrial inner membrane is required for normal growth of *Escherichia coli*  
 A:Reference number: S51801; MUID:95129546; PMID:7828591  
 A:Accession: S51801  
 A:Molecule type: DNA  
 A:Residues: 1-27, 'RNHS', 28-690 <LEI>  
 A:Cross-references: EMBL:X82612; NID:G575392; PIDN:CAAS7938.1; PID:G575393  
 C:Genetic:  
 A:Gene: SGD:ATM1; MDY  
 A:Cross-references: SGD:S0004916; MIPS:YMR301C  
 A:Map position: 13R  
 A:Genome: nuclear  
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology  
 C:Keywords: ATP; mitochondrial inner membrane; mitochondrion; nucleotide binding; P-loop  
 F:111-127/Domain: transmembrane #status predicted <TM1>  
 F:148-164/Domain: transmembrane #status predicted <TM2>  
 F:233-249/Domain: transmembrane #status predicted <TM3>  
 F:258-274/Domain: transmembrane #status predicted <TM4>  
 F:452-448/Domain: ATP-binding cassette homology <AAC>  
 F:469-476/Region: nucleotide-binding motif A (P-loop)

## Alignment Scores:

| Pred. No.:             | 5,85e-102 | Length:       | 690 |
|------------------------|-----------|---------------|-----|
| Score:                 | 1516.50   | Matches:      | 304 |
| Percent Similarity:    | 67.83%    | Conservative: | 122 |
| Best Local Similarity: | 48.41%    | Mismatches:   | 169 |
| Query Match:           | 35.65%    | Indels:       | 33  |
| DB:                    | 2         | Gaps:         | 7   |

AF133659 (1-2345) x S54211 (1-690)

```

QY 355 AAAATCATAAAGCAATGCTTTCTTATGTGTGCCCAAGACAGCCAGATCTACGAGCT 414
    |||:::|||||
Db 90 LysileuLysAspLeuPheArgTriletrProlYsgLysAenLysValArgIle 109
    |||:::|||||
QY 415 AGAGTGGCATTTGGCTGGAGATTTTGGGTGGTCAAGGCCATGATATTTGGTCCC 474
    |||:::|||||
Db 110 ArgValIleuIleAlaLeuGlyLeuLeuIleSerAlaLysIleuAenValGlnValPro 129
    |||:::|||||
QY 475 TTCATGTTTAAATATGCTGTATGACAGCTCAACAGATGTCCGGAACATGCTCAACTG 534
    |||:::|||||
Db 130 PhePhePheLysGlnThrIleAspSer-----MetAsnIle 141
    |||:::|||||
QY 535 AGTGATGACCAAAATACAGTTGACACATGGCAACA-----GCAGTTGATGATGCGTAT 588
    |||:::|||||
Db 142 AlaIleAspAspProthrIleAlaLeuProAlaIleGlyLeuThrIleLeuCytTr 161
    |||:::|||||
QY 589 GGTGTATCAAGAGCTGAGAGCTCTTTTAAACAAGTTGAAATGCAAGTATTTGGCAAG 648
    |||:::|||||
Db 162 GlyValAlaArgPheGlySerValLeuPheGlyIleuLeuArgAenAlaValPheAlaLys 181
    |||:::|||||
QY 649 GTAGCCAGATTCATCAATCCGAGAAATAGCCAAATATGCTTCCATCTTCAACAACCTG 708
    |||:::|||||
Db 182 ValAlaIleAsnAlaIleArgThrValSerLeuGlnThrPheGlnIleuMetLysLeu 201
    |||:::|||||
QY 709 GATCTGGATTTTCACTGAGCAGACAGCGAGCTTATCTAAGCTATTTGACAGAGA 768
    |||:::|||||
Db 202 AspleuGlyTrpHisLeuSerArgGlnThrGlyGlyLeuThrArgAlaMetAspArgGly 221
    |||:::|||||
QY 769 ACAAGGGATACGTTTGTCTGAGTGTGGTATTTAACTCTTCCATCATGTTT 828
    |||:::|||||

```

```

Db 222 ThrLysGlyIleSerGlnValLeuThrAlaMetValPheHisIleIlePheIleSerPhe 241
    |||:::|||||
QY 829 GAAGTCATGCTTGTGACGTGGTGTGTTTATTAACAATGCGGTGCCAGTTGCTTGGTA 888
    |||:::|||||
Db 242 GlnIleSerValValCysGlyIleLeuThrTrpGlnPheGlyAlaSerPheAlaIle 261
    |||:::|||||
QY 889 ACCCTTGAAACCTTGTATACATACACAGATTCACAGTTGACGACGCTGGAGACT 948
    |||:::|||||
Db 262 ThrPheSerThrMetLeuLeuTrpSerIlePheThrIleLysThrThrAlaTrpArgThr 281
    |||:::|||||
QY 949 AGATTGTGAATGAAGAACAACCAAGCATTAATGATGAGTAATGCTGTATGACTCA 1008
    |||:::|||||
Db 282 HisPheArgArgAspAlaAsnLysAlaAspAsnLysAlaAsnValAlaLeuAspSer 301
    |||:::|||||
QY 1009 CTGTGTAATATGAACCTGGAAGTATTTAATTAATAAAGTATGGAAGCAGACAGATAT 1068
    |||:::|||||
Db 302 LeuIleAsnPheGlnAlaValLysTrpPheAsnAsnLysLysTrpLeuAlaAspLysTr 321
    |||:::|||||
QY 1069 GATGATTTTGAAGACGTATGAGACTGCTTCATGGAAGATCCTTACTGCTATG 1128
    |||:::|||||
Db 322 AsnGlySerLeuMetAenTrpArgAspSerGlnIleLysValSerGlnSerLeuAlaPhe 341
    |||:::|||||
QY 1129 CTGACCTTGTGCAAGTCTATTTTCAAGTGTGCTTAAACAGTATTAATGCTCTGCC 1188
    |||:::|||||
Db 342 LeuAsnSerGlyGlnAenLeuIlePheThrAlaLeuThrAlaMetMetTrpMetGly 361
    |||:::|||||
QY 1189 AGTCAGGAATTTGGCAGGACCTTACTGTTGAGATCTAGTAATGGAATGAGACTG 1248
    |||:::|||||
Db 362 CysThrGlyValIleGlyLysAenLeuThrValGlyAspLeuValLeuIleAsnGlnLeu 381
    |||:::|||||
QY 1249 CTTTTCAGCTTTCATTAACCTGTAACCTTCTGGAACCTGATATATGAGACTAGACAA 1308
    |||:::|||||
Db 382 ValPheGlnLeuSerValProLeuAenPheLeuGlySerValTrpArgAspLeuLysGln 401
    |||:::|||||
QY 1309 GCACCTATATATGAACACCTGTTTACTCTACTCAAGGTAGACACCAATTAAGAC 1368
    |||:::|||||
Db 402 SerLeuIleAspMetGlnThrLeuPheLysLeuArgLysAsnGlnValLysIleLysAsn 421
    |||:::|||||
QY 1369 AAGATGATGCACTCTCCCTTCAGATCACAACGACGACT-----ACGTTGCCCTT 1422
    |||:::|||||
Db 422 -----AlaGlnArgProLeuMetLeu--ProGlnAsnValProTrpAspIleThrPhe 438
    |||:::|||||
QY 1423 GATTAATGCTATTTGAATATCATTTGAGGGCCGGAAGTCTTATGTAATATCTTTGAA 1482
    |||:::|||||
Db 439 GlnAsnValThrPheGlyTrpHisProAspArgLysIleLeuLysAenAlaSerPheThr 458
    |||:::|||||
QY 1483 GTCCCTGCAGAAAGAAAGTGGCATTGTGAGAGTGTGGTGTGAGGAAAGACAAATA 1542
    |||:::|||||
Db 459 IleProAlaGlyTrpLysThrAlaIleValGlySerSerGlySerGlySerThrIle 478
    |||:::|||||
QY 1543 GTGAGCTATTAATTTGCTTCTATGAGCCCTCAAAAGGTATGATTTATCTTGTGTC 1602
    |||:::|||||
Db 479 LeuLysLeuValPheArgPheTrpAspProGlnSerGlyLysGlyLeuIleAsnGlyArg 498
    |||:::|||||
QY 1603 AATATTAAGATGTGAGCCCTGGAAGCCTTGGAGGGCAGTGGAGTGTACTCAGAT 1662
    |||:::|||||
Db 499 AspIleLysGlnLysTrpAspIleAspAlaLeuArgLysValIleGlyAlaValProGlnAsp 518
    |||:::|||||
QY 1663 GCTGTCTCTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1722
    |||:::|||||
Db 519 ThrProLeuPheAsnAspThrIleTrpGlnAsnValLysPheGlyArgIleAspAlaThr 538
    |||:::|||||
QY 1723 CTGAGGAAGTGTATGACAGTGGCAAAATTAAGCTGACCTCATATGCAATTTCTCAAG 1782
    |||:::|||||
Db 539 AspGlnGlnValIleThrValValGlnLysValGlnLeuAlaProLeuIleLysLysLeu 558
    |||:::|||||
QY 1783 CCACATGATATGACCAACCAAGTATGGGAAACGAGACTCAAGCTTTTCAAGAGGAGAAAG 1842
    |||:::|||||
Db 559 ProGlnGlyPheAspThrIleValGlyLysArgGlyLeuMetCileSerGlyGlyLys 578
    |||:::|||||
QY 1843 CAAGAGTACCAATTCGAAGAGCCATTTTGAAGAGACCCCACTCATACTATATGATGA 1902
    |||:::|||||
Db 579 GlnArgLeuAlaIleAlaArgValLeuLeuLysAenAlaArgIleMetPhePheAspGln 598
    |||:::|||||

```







Db 463 AspThrValLeuPheAenAspThrLeuArgTyrAsnIleGlyTyrGlyLeuThrAspSer 482  
Qy 1720 TCACCTGAGGAAGTGTATGACGTGGCAAAATTAAGCTGATTCATGATGCAATTCCTTGA 1779  
Db 483 ThrAspGluGluIleGluArgAlaIleAlaLeuAlaHisIleHisGluPheIleIleSer 502  
Qy 1780 ATGCACATGTGATGTGACCCCAAGTATGGGAGACAGACATCCAGCTTCACGAGGAA 1839  
Db 503 LeuProAspGlyTyrGluThrArgValGlyIuArgGlyLeuLeuSerGlyGlyIu 522  
Qy 1840 AAGCAAGATGACAAATTTGCAAGAGCCATTTTGAAGAGCCCACTGATCTCATATG 1899  
Db 523 LysGlnArgValAlaIleAlaArgThrIleLeuLysProSerIleLeuValPheAsp 542  
Qy 1900 GAACCTATCTCATGCTTATGATTCATTAATGACAGACATATCTTGTGTCATGAGAT 1959  
Db 543 GluAlaThrSerAlaLeuAspThrHisThrGluArgGluIleGlnSerHisLeuArgGlu 562  
Qy 1960 GTGGTCAAAACACAGACTTCATATTTTCATTTGCACACAGATTTGCAACAGTGTGATGCA 2019  
Db 563 ValSerArgAspHisThrThrLeuValIleAlaHisArgLeuSerThrIleIleAspAla 582  
Qy 2020 GATGAATCATGTCTTGGATCAGAGGTAGAGCCGACGCTGATCCACATGCTTGG 2079  
Db 583 AspGluIleIleValLeuGluAlaGlySerIleValGluArgGlyArgHisAspGluLeu 602  
Qy 2080 CTTCCTAACCTCATAGTATCTATTCATTCAGAAATGTGCATACAG 2124  
Db 603 LeuLeuLys--AenGlyArgTyrAlaGlyMetTrpGlnAenGln 616  
RESULT 6  
G71731  
mitochondrial transporter atml precursor (atml) RP205 - Rickettsia prowazekii  
C/Species: Rickettsia prowazekii  
C/Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C/Accession: G71731  
R/Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U  
Nature 396, 133-140, 1998  
A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A/Reference number: A71630; MUID:99039499; PMID:9823893  
A/Accession: G71731  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-609 <RAND>  
A/Cross-references: GB:AJ235270; GB:AJ235269; NID:93860572; PIDN:CA14670.1; PID:ej134251  
A/Experimental source: strain Madrid E  
C/Genetic:  
A/Gene: atml; RP205  
C/Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology  
C/Keywords: ATP; mitochondrion  
F/372-566/Domain: ATP-binding cassette homology <ABC>  
Alignment Scores:  
Pred. No.: 6.83e-97 Length: 609  
Percent: 1446.50 Matches: 294  
Similarity: 68.67% Conservative: 118  
Best Local Similarity: 49.00% Mismatches: 170  
Query Match: 34.00% Indels: 18  
DB: 2 Gaps: 6  
AF133659 (1-2345) x G71731 (1-609)  
Qy 370 ATGCTTCTATGTGTGGCCAAAGACAGACAGATCTAGAGCTAGATGGCATTTG 429  
Db 23 LeuLeuThrTyrIleuHisPheSerLysAspPhe---AspIleArgLeuAlaGlyIleValThrSer 41  
Qy 430 CTGGAGATTTTGGTGTGCAAAAGCCATGAATATTTGGTGTCTTCATGATTTAAATAT 489  
Db 42 LeuIleCysLeuValIleAlaLysValIleAsnIlePheValProIleValTyrLysTyr 61  
Qy 490 GGTGTAGACGCTCAACGATGTGTGGGAAACATGCTGAACCTGATGATGACCAAT 549  
Db 62 IleIleAspGlyLeuAenGln-----AsnLeuSer----- 71

Qy 550 ACAGTTCGAACATGCGCAACAGAGTTCTGATTTGGCTATGCTGATTCACAGCTGAGCT 609  
Db 72 -----LeuSerValLeuIleGlyValIleIleGlyTyrGlyGlyThrLysIleLeuAla 89  
Qy 610 GCTTTTTCACAGAGTTCGAATTCAGATATTTGGACAGGTAGCCCAATTCATTCGA 669  
Db 90 GlnIlePheSerGluLeuArgValIleIlePheSerLysValGlyCysGlnAlaThrArg 109  
Qy 670 AGAATAGCCCAAAATGCTTCTTCATCTTCACATCTTCACACACCTGATCGGGATTTTCACCTGAGC 729  
Db 110 LeuValAlaLeuAsnValIlePheLysHisMetHisAsnLeuSerMetArgPheHisIleThr 129  
Qy 730 AGACAGCGGAGACCTTATATGACCTATTCAGACAGAGAAACAAGGGTATCAGTTTGTG 789  
Db 130 ArgLysThrGlyGlyLeuSerArgSerIleGluArgGlyThrLysGlyIleGluAlaVal 149  
Qy 790 CTGAGTGCCTTGGTATTTAATCTCTTCCATCATGTTTGAAGTATGATGCTTTCAGTGT 849  
Db 150 LeuArgTyrSerLeuPheAsnIlePheProThrSerLeuGluIleIleLeuValIleGly 169  
Qy 850 GTTTGTATTCACAAATGCGGTGCCAGTTGCTTGGTAACTTCGGAACATCTGTGACA 909  
Db 170 IleLeuTrpTyrPheHisGlyIleTrpPheAlaValThrLeuLeuIleThrMetIleVal 189  
Qy 910 TACACAGCATTCACAGTTGCAGTCACACGCGTGAGACCTAGATTTGAATGAATGAAAC 969  
Db 190 TyrValCysTyrThrLeuLeuIleSerThrTrpArgIleSerPheAlaArgGluMetAen 209  
Qy 970 AAAGCATATATGATGACAGGTAAATGCTGCTATAGACTCACTGCTGAATTAAGAACTGTG 1029  
Db 210 GlnSerAspAsnThrAlaAsnAsnArgAlaIleAspSerLeuLeuAsnPheGluThrVal 229  
Qy 1030 AAGTATTTTAAATATGAAGATATGAAGACACAGATATGATGATTTTGAAGACGTAT 1089  
Db 230 LysTyrPheAsnAsnGluGluTyrGluAlaIleLysPheAsnAspAlaLeuGlnAlaTyr 249  
Qy 1090 GAGACGCTTCATTCAGAAAGTAACTCTACTGCTGCTATGCTGAACCTTGGTCAAGTGT 1149  
Db 250 GluLysSerAlaThrLysIleThrAsnSerLeuSerIleLeuAsnIleGlyGlnAspVal 269  
Qy 1150 ATTTTACAGTTCGCTTTAACACGCTATATATGTCCTGCCACATGACGGAATTTGGCCAGGT 1209  
Db 270 IleIleSerLeuGlyLeuValSerLeuMetIleLeuSerValAsnAlaIleAenGlnAen 289  
Qy 1210 ACCCTTACTGTTGAGATCTAGTAATGAGATGAGATGACCTGTTTTTACGTTTATTAACC 1269  
Db 290 LysMetMetValGlyAspLeuIleMetValAsnAlaTyrLeuPheGlnLeuSerIlePro 309  
Qy 1270 CTGAACCTTCTGGGAACCTGATATATAGAGACTAGACACAGCACTCATATAGAACACC 1329  
Db 310 LeuSerIleLeuGlyPheAlaTyrArgGluIleLysAsnAlaLeuValAsnMetGluAsp 329  
Qy 1330 TTGTTTACTCTACTCAAGGTAGACACCAATTTAAAGACAAAGTATGATGATCTCCCTT 1389  
Db 330 MetPheLysLeuLeuAspIleProAlaGluIleGlnAspSerValAspAlaLysGluLeu 349  
Qy 1390 CAGATCAACACACAGACACTACCGGCTGCTTGAATATGTCATTTTGAATTCATTTGAG 1449  
Db 350 IleIleLeuSer-----LysCysLysValSerPheAsnAsnValSerPheAlaTyrAenLys 367  
Qy 1450 GGCCAGAAAGTCCCTTATGTAATATCTTTGAAGCTCCGACAGAAAGAAAGGAGCCATT 1509  
Db 368 GluArgThrIleLeuHisAsnIleThrPheThrIleGluSerCylLysThrIleAlaVal 387  
Qy 1510 GTAGAGGTAGTGTGACGAGGAAAGACAAATATGAGGCTATATTTGCTTCTTATGAG 1569  
Db 388 ValGlySerSerGlyAlaIleLysSerThrIleSerArgLeuLeuPheArgPheTyrAsp 407  
Qy 1570 CCTCAAAAGGTATGATATCTGCTGCTCAAAATATACAGATATGACCTGGAAGAC 1629  
Db 408 IleAsnSerGlySerIleIleAspAsnGlnAspIleArgGluValLysGlnGlySer 427

```

QY 1630 CTTCGAGCGAGTGGAGTGTACCTCAGAGTGTCTCTTCATATAATCTATTAT 1689
Db LeuArgLysSerIleGlyIleValProGlnAspThrValLeuPheAsnAspThrIleTyr 447
QY 1690 TACAACTCTTATATGAAACATCACTGCTTCACTGAGAGTATGACATGCGCAAA 1749
Db TyraAsnIleAlaTyrGlyAsnAsnAlaAspTyrAspGluValIleAlaAlaAspLys 467
QY 1750 TTACGTGACCTTCATGATGATGATTCCTGATGATGATGATGATGATGATGATG 1809
Db AsnAlaHisIleHisGluPheIleSerValLeuProGluGlyTyrAlaThrGlnValGly 487
QY 1810 GAACGAGGACTTCAAGCTTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1869
Db GluArgGlyLeuLysLeuSerGlyGlyGluSerGlnArgIleAlaIleAlaArgThrIle 507
QY 1870 TTGAAGAGACCCCGCTCATCTACTATGATGATGATGATGATGATGATGATGAT 1929
Db LeuLysAsnProSerIleTyrValPheAspGluAlaThrSerSerLeuAspThrLysThr 527
QY 1930 GAAGAGACTATCTTGTGGCCATGAGAGATGAGTCAACAGAGACTTCTATTTCAT 1989
Db GluLysLeuIleGlnAlaSerLeuLysGluIleSerAlaAsnHisThrThrLeuIleIle 547
QY 1990 GCACACAGATTGTCAACAGTGTGTATGATGATGATGATGATGATGATGATGAT 2049
Db AlaHisArgLeuSerThrIleValAspAlaAspGluIleIleValLeuAspAsnGlyTyr 567
QY 2050 GTACCCGAACGTGTGATCCCATCATGTTGCTGCTTAACCTCATATGATCTATTCAG 2109
Db 568 IleValGluArgGlyAsnHisLysThrLeuLysAsn--GlnGlyTyrTyrAlaGlu 586
QY 2110 ATGTGCTATACACAGAGCGGTGTGCGAGACATGATTAACCCCAATGGGAGCAA 2167
Db 587 LeuTyrTyrLysGlnGlnGlnGluCys--AsnGlnAsp-ThrGlnAsnCysLysGln 604

```

RESULT 7  
AF3438 heavy metal tolerance protein precursor [imported] - *Brucella melitensis* (strain 16M)  
C:Species: *Brucella melitensis*  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 17-May-2002  
C:Accession: AF3438

R:DelVecchio, V.G.; Kapurali, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, M.; Mazur, M.; Golezman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Leles, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AF3438  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-628 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AAL52673.1; PID:917983498; GSPDB:GN00190  
A:Experimental source: Strain 16M  
C:Genetics:  
A:Gene: BME11492  
A:Map position: I  
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Alignment Scores:  
Pred. NO.: 1.61e-93 Length: 628  
Score: 1400.00 Matches: 284  
Percent Similarity: 64.66% Conservative: 122  
Best Local Similarity: 45.37% Mismatches: 192  
Query Match: 32.91% Indels: 28  
DB: 2 Gaps: 6

AF133659 (1-2345) x AF3438 (1-628)

```

QY 346 GATGATCGGAAATATCAATAAAGCATGCTTTCTATGATGCGCAAGAGAGCCAGAT 405
Db 13 GuitrLLeuLysThrLeuArgAsnLeuTyrProTyrMetTyrProSerAspArgProAsp 32
QY 406 CTACAGAGCTAGAGTTGCCATTTTCGTGGGATTTTGGGTGTGCAAGGCCATGAATATT 465

```

```

Db 33 LeuArgMetArgValIleTyrAlaThrPheTyrLeuValLeuSerLysIleValLeuIle 52
QY 466 GTGGTCCCTTCATGCTTAAATATGCTGTAGACAGCCCTCAACAGATGCGGAAACATG 525
Db 53 LeuValProTyrPhePheLysTyrValThrAsnAlaLeuAsn-----GlyGlnLeu 69
QY 526 CTGAACCTGACTGATGCAACCAATATACAGTTGCACATG-----GCAACGACAGTT 576
Db 70 -----HisAlaProTyrTyrIleProValValLeuValGlyAlaValMetLeu 85
QY 577 CTGATGGCTATGCTGTATGATCAAGAGCTGAGCTCTTTTAAACGAACTTGCAGCA 636
Db 86 ValLeuAlaTyrAsnAlaAlaLysIleValGlnAlaGlyLeuAsnGlnLeuArgAspAla 105
QY 637 GTATTGGCAAGTGAAGCCCAATTCATCGAAGATGACCAAAATGCTTCTTCAT 696
Db 106 LeuPheAlaSerValGlyGlyTyrAlaValArgGlnLeuAlaTyrLysThrPheValHis 125
QY 697 CTTCACAACTGATGATGCTGTTTTCACCTGAGCAGACAGACGAGCTTATCTTAAGGCT 756
Db 126 MetHisGluLeuSerLeuArgPheHisLeuGlnArgThrGlyGlyLeuSerArgVal 145
QY 757 ATTGACAGAGGAGAACAGGGGTATCAGTTTGTCTGAGTGTCTTGATTTATCTTCTT 816
Db 146 IleGluArgGlyThrLysGlyIleGluThrIleValArgPheThrIleLeuAsnThrLeu 165
QY 817 CCCATCATGTTTGAAGTGAAGTGTGTGAGTGTCTTGTGATCAAAATGCGGCCAG 876
Db 166 ProThrIleLeuGluPheAlaLeuThrAlaValIlePheAlaPheAlaTyrGlyLeuSer 185
QY 877 TTTGCTTTGGTAACCTTGGAACTTGTGATACATACACAGATTCACAGTTGACAGTACA 936
Db 186 TyrLeuValValIleAlaIleThrValTyrLeuTyrThrThrPheThrIleArgAlaSer 205
QY 937 CCGTGGAGACTAGATTAGATTTGAAATGAAATGAAACAGATTAATGATGACGTTAATGCT 996
Db 206 AspTyrArgIleAsnIleArgArgGluMetAsnAspSerAspThrAspAlaAsnThrLys 225
QY 997 GCTATGACCTGACGCTGGAATATGAAACGTGAAAGTATTTATATGAAAGATATGAA 1056
Db 226 AlaIleAspSerLeuLeuAsnPheGluThrValLysTyrPheGlyAsnGluAlaMetGlu 245
QY 1057 GCACAGAGATATGATGATTTTGAACAGTATGAGACGTGCTTCATTTGAAAGTACCTCT 1116
Db 246 AlaLysArgPheAspGlyAlaMetAlaArgTyrGluLysAlaIleThrGlnThrTyrThr 265
QY 1117 ACTGTGCTATGCTGAACCTTGTGCAAGTGTCTATTTTCACTGTGCTTTAACAGCTATA 1176
Db 266 SerLeuGlyTyrLeuAsnPheGlyGlnAlaValIlePheGlyAlaGlyMetAlaIleVal 285
QY 1177 ATGTGCTCGGCACATGACGGAATGTGTGCGAGTACCTTACTGTTGAGATCTATGATG 1236
Db 286 MetValMetSerAlaMetGluValGlnLysGlyThrGlnSerLeuGlyAspPheValPhe 305
QY 1237 GTGATGAGACTGCTTTTTCAGCTTTCATTAACCTTGAACCTTTCGTGGAACTGTATATGA 1296
Db 306 IleAsnAlaLeuLeuMetGlnLeuSerIleProLeuAsnPheIleGlyPheIleTyrArg 325
QY 1297 GAGACTAGACAGACCTCATATGATATGAACACTTGTTTACTTACTTCAAGTACAGACACC 1356
Db 326 GluIleArgGlnGlyLeuThrAspIleGluGlnMetCpheaPheLeuLeuAspValLysGln 345
QY 1357 CAATTTAAAGCAAGATGATGCGATCTCCCTTCAAGATCAACACAGACAGCTACCGTG 1416
Db 346 GluValLysAspLysProGlyAlaProAlaLeuLysVal-----AspSerGlyAlaIle 363
QY 1417 GCTTTGATATGTCATTTTGAATACATTTGAGGCGCAAGAGTCTTATGTAATATCC 1476
Db 364 SerPheLysAspValHisPheAlaTyrAspProGlnArgProIleLeuArgGlyIleSer 383
QY 1477 TTTGAAGTCCCTGACAGAGAAAGATGGCCATTGTAGAGGACTAGTGGTCAAGGAAAGC 1536

```



|    |      |  |      |
|----|------|--|------|
| Db | 321  | IIeqlgYpHevVallYrAqGuIlIeAqGInGlyLeuThrAspIleGlulMetHePheAsp     | 340  |
| Qy | 1339 | CTTACTCAGGTAGACACCCAAATTTAAAGCAAAAGTATGGCATCTCCCTTCAGATCACA      | 1398 |
| Db | 341  | LeuEugIuValIGuIaIGuIValThrAspArgProAspAlaIySerProIeuuIaIaIGly    | 360  |
| Qy | 1399 | CCACAGACAGCTACCGCGCTTGATATATGTCATTTTGAAATCACTTGAGGCGCAGAA        | 1458 |
| Db | 361  | Pro-----GlyAlaIleSerPheAspValHisPheAlaIyrAspProIuIuArgPro        | 378  |
| Qy | 1459 | GTCCTTAGTGAATATCTTTGAAAGTCCCGCAGAGAAAGAAAGTGCCCTTGTAGAGAT        | 1518 |
| Db | 379  | IleIeuYsIGlyValSerPheAspValProIaIGlyLeuThrValAlaIleValIGlyPro    | 398  |
| Qy | 1519 | AGTGGGTGAGGAAAGACACAAATAGTGAAGCTATTTATTTGCTTCTATGAGCTCAAG        | 1578 |
| Db | 399  | SerIGlyAlaIGlySerThrIleSerArgIleuLeuYrArgPheYrAspIleGIInGlu      | 418  |
| Qy | 1579 | GGTAGCAATTATCTGTGTCGCAAAATATTCAGAGTGAAGCTGGAAAGCTTGAGAG          | 1638 |
| Db | 419  | GlyAlaValThrIleAspIGlyGlnAspIleArgAspValThrGlnYsSerIeuArgSer     | 438  |
| Qy | 1639 | GCAATGGAGAGTGTACTCTCAGATGCGTCCTCTTCATATATCATTTATTTACAACTC        | 1698 |
| Db | 439  | MetIleGlyMetValProGlnAspThrValLeuPheAsnAspThrIeuuIaIyrAsnIle     | 458  |
| Qy | 1699 | TTATATGGAAACATCAGTCTTCACTCGAGAAAGTATAGCAGTGGCAAAATTAGCTGA        | 1758 |
| Db | 459  | ArgYrIGlyArgProSerAlaThrAspGluIleuYsIleAlaIaIaAspAlaIeGln        | 478  |
| Qy | 1759 | CTTCATGATGCAATCTTCGAAAGCCCATGATATGACACCCAGTAGGAGACAGAGA          | 1818 |
| Db | 479  | IleSerAlaPheIleIGlyYsIeuProAspGlyTyrAlaIleuMetValIGlyIuArgGly    | 498  |
| Qy | 1819 | CTCAAGCTTTCAGAGAGAGAAAGCAAAAGTAGCAATTGCAAGACCTTTTGAAGAC          | 1878 |
| Db | 499  | LeuYsIeuSerIGlyIGlyIuYsGlnArgValAlaIleAlaYrThrIleuYsAla          | 518  |
| Qy | 1879 | CCCCAGTCATACCTATGATGGAAGCTACTTCATCGTTGATTTCCATTCTGAAGACT         | 1938 |
| Db | 519  | ProProIleIeuIleIeuAspGluAlaIleThrSerAlaIeuYsPthrIleThrIGuIInGlu  | 538  |
| Qy | 1939 | ATTCTGGTGTGAGAGAGATGTGGTCAAAACAGAGAACTTCTTATTTCCATTGCACAGA       | 1998 |
| Db | 539  | IleGlnSerAlaIeuAspIleValaSerYsAsnArgThrThrIeuValIleAlaIleArg     | 558  |
| Qy | 1999 | TTGTCAACAGTGTGTGATGCAAGATGAATATATTGTCTTGATCAGGGTAAAGTACCGAA      | 2058 |
| Db | 559  | LeuSerThrValIleHisAlaAspGluIleIleValIleuYsIGuIuYsIleuIleAGlu     | 578  |
| Qy | 2059 | CGTGTATCCCAACCAATGGTGTGCTGTGCTAACCCCTCATATGATCTATTCACAAATGTGGAGT | 2118 |
| Db | 579  | ArgGlyIleHisAlaSerIeuMetAlaHis--AspGlyLeuYrAlaSerMetIrrPser      | 597  |
| Qy | 2119 | ACACAGACAGCCCGTGTGCAGAAACCATGATTAACCCAAATGGGAGACAAAGAAAT         | 2178 |
| Db | 598  | ArgIle-----ArgIuIuIa602  |      |
| Qy | 2179 | ATATCCAAAGAGAGAAAGAAAGAAATCTACAAAGAA2214                         |      |
| Db | 603  | IleArgIaIGluIGuIuMetIeuArgHisValArgIuArgIu614                    |      |

RESULT 9  
H97489  
mitochondrial transporter atmi precursor (atmi) rp205 [imported] - Agrobacterium tumefaciens  
C/Species: Agrobacterium tumefaciens  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C/Accession: H97489  
R/Goodner, B.; Hinkle, G.; Gatung, S.; Miller, N.; Blanchard, M.; Quirollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.  
Science 294, 2223-2328, 2001  
Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens

A:Reference number: A97359; PMID:11743194  
A:Accession: H97489  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-629 <KOR>  
A:Cross-references: GB:AE007869; PDB:AAK6873.1, PDB:5J556091; GSPDB:GN00165  
C:Genetics:  
A:Gene: AGR\_C\_1966  
A:Map position: circular chromosome

|                        |          |
|------------------------|----------|
| Alignment Scores:      |          |
| Pred. No.:             | 1.04e-91 |
| Score:                 | 1375.00  |
| Percent Similarity:    | 65.36%   |
| Best Local Similarity: | 45.92%   |
| Query Match:           | 32.32%   |
| DB:                    | 2        |
| Length:                | 629      |
| Matches:               | 281      |
| Conservative:          | 119      |
| Mismatches:            | 190      |
| Indels:                | 22       |
| Gaps:                  | 4        |

AF133659 (1-2345) X H97489 (1-6229)

[illegible]

```

Db      282  IleglyserThrIleMetMetValMetSerIleAlaValAlaGlnArgGlnGlnThr 301
      1219  GTTGAGAGTACTAGTAATGTAATGAGCTGCTTTTTCAGTTTTCATTAACCCCTGAACCTTT
      302  IleglyAspPheValAlaPheValAlaPheValAlaPheValAlaPheValAlaPhe 321
      1279  CTGGGAAGTGTATATAGAGACATGACACAGACTCATAGATATGACACCTGTGTTACT 1338
      322  IleglyPheValTyrArgGlnIleArgGlnIleArgGlnIleArgGlnIleArgGln 341
      1339  CTACCAAGGTAGAGACACCAATTAAGACAAAGATGAGATGCTCCCTTCATATGACA 1398
      342  LeuLeuGlnValAlaGlnValAlaThrAspArgProAspAlaValProLeuAlaGly 361
      1399  CCACAGACAGCTACCGCTTGTATGATGCAATTTGATGCAATTTGAGGCGCAGAAA 1458
      362  Pro-----GlyAlaIleSerPheArgAspValIleSphAlaTyrAspProGlnArgPro 379
      1459  GTCTTATGTAATATCTTTGAAAGTCCCTGACGAAAGAAAGTGGCCATTGAGAGGT 1518
      380  IleLeuysGlyValSerPheAspValProIleGlySerThrValAlaIleValGlyPro 399
      1519  AGTGGTCAAGGAAAGACAAATAGTAGGCTATTTGCTTCTATGAGCTCAAAAG 1578
      400  SerGlyAlaGlySerThrIleSerArgLeuLeuTyrArgPheTyrAspIleGlnGln 419
      1579  GGTAGCAATTTATCTGCTGCAAAATATACAGATGAGCTGGAAGGCTTGGAGG 1638
      420  GlyAlaValThrIleAspGlyGlnAspIleArgAspValThrGlnValSerLeuArgSer 439
      1639  GCATGGAGAGTGTACTCAGAGATGCTGCTCTTCATATATCTATTTTCAACCTC 1698
      440  MetIleGlyMetValProGlnAspThrValLeuPheAsnAspThrLeuAlaTyrAsnIle 459
      1699  TTATATGGAACATCAGTGTCTTCACTGAGAAATGTATGAGAGGCAAAATTTAGCTGGA 1758
      460  ArgTyrIleArgProSerAlaThrAspGlnGlnLeuysAlaAlaAspAlaIleGln 479
      1759  CTTTCATGATCAATTTCTCGATGCCACATGATATGACACCCAGTAGGGAGACGAGA 1818
      480  IleSerIleAlaPheIleGlySerLeuProAspGlyTyrAlaThrMetValGlyGlnArgGly 499
      1819  CTCAAGCTTTCAAGAGAGAAAGCAAAAGTAGCAATTCGACAGAGCATTTTGAAGAC 1878
      500  LeuysLeuSerGlyGlyGlnValArgValAlaIleAlaArgThrIleLeuysAla 519
      1879  CCCCAGTCACTCTATGATGAAAGTACTTATCTGATGATTCGATTAAGAGACT 1938
      520  ProIleIleLeuIleLeuAspGlnAlaThrSerAlaLeuAspThrIleThrGlnGln 539
      1939  ATTCTTGATGCGATGAGATGTGTCAACACAGAACTTATTTTCATTTGACACAGAGA 1998
      540  IleIleIleIleAlaLeuAspIleValSerIleAsnArgThrIleLeuValIleAlaIleAsnArg 559
      1999  TTGTCAACAGTGTGTGATGACATGAATGATGTCTTGTGATCAGGTAAAGTAGCCGAA 2058
      560  LeuSerThrValIleIleIleAlaAspGlnIleIleValLeuysGlnGlyLeuIleAlaGln 579
      2059  CGTGTATCCCAACATGCTTTGCTTGTCAACCTCATAGTATCTATTCAGAAATGTGGCAT 2118
      580  ArgGlyThrIleAlaSerLeuMetAlaIleAspGlyLeuTyrAlaSerMetTyrSer 598
      2119  ACACAGACGACCGGTGTGCAGAACCATGATTAACCCAAATGGGAGAGAAAGAAAT 2178
      599  ArgGln-----ArgGlnAla 603
      2179  ATATCCAAAGAGAGAGAAAGAAAGAAATCAAGAA 2214
      604  IleArgIleGlnGlnMetLeuArgHisValAlaArgIle 615

```

RESULT 10  
JB0248

```

ATP-binding cassette half-transporter - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 05-Dec-1998 #sequence_rev15ion 05-Dec-1998 #ext_change 21-Jul-2000
C/Accession: J02048
R/Hirsch-Ernst, K.I.; Gairi-Rahimi, S.; Ernst, B.P.; Schmitz-Saue, C.; Blume, S.; Kahl
Biochem. Biophys. Res. Commun. 249, 151-155, 1998
A/Title: Molecular cloning and tissue distribution of mRNA encoding a novel ATP-bi
A/Reference number: J02048; M01D:98381042; EMD:9705847
A/Accession: J02048
A/Molecule type: mRNA
A/Residues: 1-836 <HIR>
A/Cross-references: GB:AJ003004; NID:g2970020; PIRN:CA05793.1; PID:g2982567
C/Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C/Keywords: ATP
F/606-800/Domain: ATP-binding cassette homology <ABC>

Alignment Scores:
Pred. No.: 2,04e-82 Length: 836
Score: 1247.00 Matches: 280
Percent Similarity: 54.94% Conservative: 126
Best Local Similarity: 37.89% Mismatches: 247
Query Match: 29.31% Indels: 86
DB: 2 Gaps: 13

AF133659 (1-2345) x JB0248 (1-836)
      28  CGCTGGGCGGCGCGGCGGCTGCTTTCGAAAGCGCGGACATCCGGATTCGATC--- 84
      133  ArgGlnSerLeuAlaMetGlyValTyrMetLysPheArgIleSerLeuGlyLeuLeu 152
      85  -----CGGCTTATGCTCTGTTTGAAGCGGCTCAGGTCCGAG 120
      153  LeuThrThrValThrPheAlaIleGlnValLeuValLeuValSerThrPheAsnSerProGln 172
      121  TGGAGGCCACATCAACTCGGCGCTTGGGAGACCGCTCGAGCTTACAGATTCAGAGTCA 180
      173  Trp----- 173
      181  TTAATAAGTATCATGCGAGATGAGATGGAAAGCAATTCAGAGATGTTAGATGCT 240
      174  -----TrpThrSerArgAlaAspLeuGlyGlnGlnValGlnPhe----- 186
      241  GCAAGGCTCTCCAGGTATGCGCATGATGAAAGAGACA----- 282
      187  -----GlyLeuThrValLeuArgTyrMetThrSerGlyLeuPheIleLeu 202
      283  ---TGTTGC----- 288
      203  GlyLeuThrAlaProGlyLeuArgProGlnSerTyrThrLeuHisValAlaGlnGlnAsp 222
      289  ---CATGTCATGTCAGAGAGAGACATCCACAGACCCAAAGAGGTTAAAGATGTT 345
      223  GlnAspGlyGlyArgAsnGlnGlyArgSerThrAspProArgSerThrTrpAlaAspLeu 242
      346  GATATCTCGAATAATCATAAAGCATGCTTTCTTATGTGTGCGCCAAAGACAGGCCAGAT 405
      243  GlyArgLys-----LeuArgLeuLeuSerGlyTyrLeuThrProArgGlySerProSer 260
      406  CTACGAGCTAGAGTTCATTTGCTGAGATTTTGGTGGTGGCAAGAGCCATGAATATT 465
      261  LeuGlnLeuThrValLeuLeuLeuCysMetGlyLeuMetGlyLeuAspArgAlaLeuAsnVal 280
      466  GGGTTCCTTCATGTTTAAATATGCTGTAGACAGCTCAACCAAGATGTCGGGAAACATG 525
      281  LeuValProIlePheTyrArgAspIleValAlaLeuLeuThr----- 294
      526  CTGAACCTGATGATGACCA-----AATACAGTTGACACCATGCAACAGAGTTCTG 579
      295  -----SerLysAlaIleProTyrSerSerLeuAlaThrThrValThrTyrValPhe 311
      580  ATTGCGTAT-----GGTATCAAGAGCTGAGCTGCTTTTATACAGAACTGGAAT 633
      312  LeuLysPheLeuGlnGlyGlyGlyThrGlySerThrGlyPheValSerAsnLeuArgThr 331

```

|    |      |  |      |
|----|------|--|------|
| Oy | 634  | GCGAATTTGGCAGGTGACCCCAAAATTCAATCCGAAGAATGCCAAAATAATGCTTTC      | 693  |
| Dd | 332  | PheLeuTrpIleArgValGlnGlnPheThrSerArgValGluGluMetArgPheSer      | 351  |
| Oy | 694  | CATCTTCAACAACCTGCATCTGGGCTTTTCACTCGACAGACAGACGGAGCTTATCTGA     | 753  |
| Dd | 352  | HisIleuHisGluLeuSerLeuArgTrpHisIleuGIlyrArgTrnGlyGluValLeuArg  | 371  |
| Oy | 754  | GCTATTGACAGAGAACAGGGGATTCAGTTTGTCCTGAGTGGCTTGGTATTAAATCTT      | 813  |
| Dd | 372  | IleValAspArgGlyThrSerSerValThrGlyLeuLeuSerTyrlleuValPheAsnIle  | 391  |
| Oy | 814  | CTTCCCACCATGTTTGAAGTATGATCTGTCAAGTGGTGGTTTTGTTATCAAAATGC-----  | 867  |
| Dd | 392  | IleProthrIleuAlaAspIleIleIle-----GlyIleIleTyrlPheSerMetPhePhe  | 409  |
| Oy | 868  | GGTGGCCAGTTGCTTGGTATGACCCTTGGAAACTCTGGATATACAGCATTCACAGTT      | 927  |
| Dd | 410  | AsnAlaTrpPheGlyLeuIleValPheLeuCyMetSerLeuTyrlleuIleThrIle      | 429  |
| Oy | 928  | GCAGTCACACGGTGGAGAACTGATTTTAGATTAAGAAATGAAACAAGACAGATATGATGCA  | 987  |
| Dd | 430  | MetValTrnGluTrpArgAlaLysPheArgArgAspMetCsnTrnGlnGluAsnAlaThr   | 449  |
| Oy | 988  | GGTATGCTGCTATAGATCTCACTGCTGATATTAATAAACTGAGAGATTTTAAATGAA      | 1044 |
| Dd | 450  | ArgAlaArgAlaValAspSerLeuLeuAsnDnheGluThrValIlysrTyrlYAsnAlaGlu | 469  |
| Oy | 1048 | AGATTTGAAGCACAGACATATGATGAGATTTTGAAGACATATGAGATGCTTCATTGAAA    | 110  |
| Dd | 470  | GlyTyrlGluLeuGluArgLysArgGluAlaIleLeuLysPheGlnGlyLeuGluTrpLys  | 489  |
| Oy | 1108 | AGTACCTTACTGTCGGCATCTCAACTTGGTCAAAAGTGCATATTTTCACTGTCGGTTA     | 116  |
| Dd | 490  | SerThrAlaSerLeuValLeuLeuAsnDnTrnGlnAsnMetValIleGlyPheGlyLeu    | 509  |
| Oy | 1168 | ACAGCTAATATGTCCTGCCAGTCAAGGAAATTGGCAGGTACCTTACTGTTCAGAT        | 122  |
| Dd | 510  | LeuAlaGlySerLeuLeuCysAlaTyrlPheValSerGluArgLysGluGlnValGlyAsp  | 529  |
| Oy | 1228 | CTAGATGAGGGGAATGAGCTGCTTTTACGCTTTCATTACCCCCGAACTTTGGGAAT       | 128  |
| Dd | 530  | PheValLeuPheGlyTrnTyrlIleThrGlnLeuTyrlMetProLeuAsnTrpPheGlyThr | 549  |
| Oy | 1288 | GTAATATAGAGAGACTGACAAAGCACTCATAGATATGAACCTTGTTACTTAATCTCAAG    | 134  |
| Dd | 550  | TyrTyrlArgMetIleGlnThrAsnPheIleAspMetGluAsnMetPheAspLeuLeuLys  | 569  |
| Oy | 1348 | GTAGACACCCCAATTAAGAACAAGTATGCGATCTCCCTTCAGATCAACACAGACA        | 140  |
| Dd | 570  | GluGluTrnGluValLysAspValProGlyAlaGlyProLeuArgPhe-----HisLys    | 587  |
| Oy | 1408 | GCTACCGTGGGCTTTGATTAATGTGCATTTTAATATCAATTTGAGGCCCAAGTCTTAGT    | 146  |
| Dd | 588  | GlyArgValGluPheGluAsnValHisPheSerTyrlAlaAspGlyArgGluThrLeuGln  | 607  |
| Oy | 1468 | GGATATATCTTTGAAGTCCCTGCACAGAAAGAAAGTGGCCATTGTAGAGGTATGGGTCA    | 1522 |
| Dd | 608  | AspValSerPheThrValMetProGlyGlnThrValAlaLeuValGlyProSerGlyAla   | 627  |
| Oy | 1528 | GGGAAACACAAATGATGAGGCTATTAATTTGGCTTATAGAGCTCAAAAGGAGCATT       | 158  |
| Dd | 628  | GlyLysSerTrnIleLeuArgLeuLeuPheArgPheTyrlAspIleSerSerGlyCysIle  | 647  |
| Oy | 1588 | TATCTTGCTGTCAAAATATATCAAGATGAGCTGGAAAAAGCTTGGAGGACGTGGA        | 1644 |
| Dd | 648  | ArgIleAspGlyGlnAspIleSerGlnValTrnGlnIleSerLeuArgSerHisIleGly   | 667  |
| Oy | 1648 | GTGGTAACCTCAGATGCTGTCTCTTCATATAATCTATTATTCACACCTTATATGAA       | 1707 |
| Dd | 668  | ValValProGlnAspThrValLeuPheAsnAspTrnIleAlaAsnAlaIleArgTyrlGly  | 687  |

|    |      |   |      |
|----|------|---|------|
| Qy | 1708 | AAAGTAAAGGCTACACCTAGAGAAAGGTATGGAGAGGCAAAATTAAGCTGACCTTATAT       | 1767 |
| Db | 688  | ArgValThrIaIeLysPserGluIleGlnAlaIaIaGlnAlaIaGlyIleHisP            | 707  |
| Qy | 1768 | GCATTTCTTGAATGCGACATGGATATGACACCCAAAGTAGGGGAACGAGATCAAGCTT        | 1827 |
| Db | 708  | AlaIleuSerPheProGluGlyTyrGluThrGlnValGlyValArgGlyLeuIleuSer       | 727  |
| Qy | 1828 | TCAGAGAGAGAAAAGCAAAAGTAGTACCAATTCGAAAGACCATTTTGAAGACCCCACTC       | 1887 |
| Db | 728  | SerItylGluIuysGlnArgValAlaIaIaAlaArgThrIleLeuIleuAlaProPheIle     | 747  |
| Qy | 1888 | ATACTCATAGTAAAGCTACTTCATGCTGGTAAATGGATTAAGTAGAGAGACTATTTCTGCT     | 1947 |
| Db | 748  | IleLeuIeuuSpGlnIaIaThrSerIaIaIeuaPhePheSerAenGluArgAlaIleGlnIa    | 767  |
| Qy | 1948 | GCCATGAGAGATGTGGTCAAAACACAGAACTTATTTTCATTCGACACAGATTTGCAACA       | 2007 |
| Db | 768  | SerIeuAlaIuysAlaCysThrAenAlaGluThrIleValaValaIaIaHisArgLeuSerThr  | 787  |
| Qy | 2008 | GTGGTGTATGACAGATGAATATCATTTGTTGGATCAAGGGTAAAGTACCCGAACSTGTACC     | 2067 |
| Db | 788  | ValValaenAlaPheGlnIleLeuValIleIleuysPglyCysIleIleIeGluArgGlyIuArg | 807  |
| Qy | 2068 | CACCATGTTTGGTCTGCTAACCTCATAGTATCTATTTCAGAAATGTGCGCATACAG          | 2124 |
| Db | 808  | HisGluIaIeLeuSerPheArg--GlyGlyValaItyrAlaGluMetTrpGlnIleuGln      | 825  |

## RESULT 11

ABC transporter, HlyB/MsbA family CCI314 [imported] Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: C87412  
R:Bernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heid  
B.; Lab, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D  
N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fr  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; M01D:21173698; PMID:11259647  
A:Accession: C87412  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-643 <STO>  
A:Cross-references: GB:AE005673; NID:gl3422655; PIDN:AAK23295.1; GSPDB:GN00148  
C:Genetic:  
C:Gene: CCI314

**Alignment Scores:**

|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 1.43e-67 | Length:       | 643 |
| Score:                 | 1042.07  | Matches:      | 293 |
| Percent Similarity:    | 52.07%   | Conservative: | 93  |
| Best local Similarity: | 37.43%   | Mismatches:   | 266 |
| Query Match:           | 24.49%   | Indels:       | 58  |
| DB:                    | 2        | Gaps:         | 7   |

AF133659 (1-2345) x C87412 (1-643)

|    |     |  |     |
|----|-----|--|-----|
| Oy | 145 | TTGGGAAACCGCTGCAGCCCTTACCAGATTCCTCAGACTCATTTAAAAAGTACACANTGCCAGAGA | 204 |
|    |     |  |     |
| Db | 5   | LengllAspAlaAargArgTlPargThrAspHisleuAlaaspProlysglyTlPAlaIeuc     | 24  |
|    |     |  |     |
| Oy | 205 | TTGGGAATAAGGCAT-----TCAGAGACAGTCTTGAT                              | 237 |
|    |     |  |     |
| Db | 25  | ArggllyLengllglylProprioserProglYArgluAlavalAlaglyAlavalAlaLys     | 44  |
| Oy | 238 | GCTGCATAAGGCTCTCCAGGTATGGCCA CTGATAGAAAAAGAGACANTGGCATGTGTCAT      | 297 |
|    |     |  |     |
| Db | 45  | AlaaspGlnProIleargspHetip-----                                     | 52  |
| Oy | 298 | GCAGAGAGAGATCTCCACACAGAACCCCAAGAGGTTAAAAGATTGTGATCTCGGAAA          | 357 |
|    |     |  |     |
| Db | 53  | -----Lys   | 53  |



|    |      |   |      |
|----|------|---|------|
| OY | 358  | ATCATTAAGCAAGCTTTCTTAATGTGTGGCCCAAGACAGGCCAGATCTACAGACTAA       | 417  |
| Db | 54   | AlaMetSerAspLeuLeuGlnLeuValLeuArgSerGlnAlaProGlyLeuArgTyrArg    | 73   |
| OY | 418  | GTTCGCACTTGTGGCTGGGATTTTGTGGGTGGCAAGGCCATGAAATATGTGGTCCCTTC     | 477  |
| Db | 74   | LeuThrValAlaLeuLeuLeuThrLeuThrGlyLysValLeuGlyValLeuAlaProLeu    | 93   |
| OY | 478  | ATGTTTAAATATGCTGTAGACAGCTCAACCCAGATGTGGGAAACATGTGCACCTGACT      | 537  |
| Db | 94   | MetLeuGlyGlnAlaVal-----AsnGlnLeuSerValGly                       | 105  |
| OY | 538  | GATGCACCAAAATACAGTTTGCACCATGGACACAGCACTCTG--ATTGGCTATGGGTGA     | 594  |
| Db | 106  | GlnGlyValAlaValAlaThrValThrLeuAlaPheAlaSerLeuAlaIleGlyTyrAlaLeu | 125  |
| OY | 595  | TCAAGAGCTGGAGCGCTTTTTTTAAAGAAAGTTGAAATGCAGATATTGTGGCAAGTACC     | 654  |
| Db | 126  | ValArgPheIleSerAlaAlaAlaProGlnAlaArgSerPheThrIlePheThrProValAla | 145  |
| OY | 655  | CAGATTTCAATCCGAGAAATATGCCAAAAATGTCTTCTCCATCTTGCACAACTGTAGTGC    | 714  |
| Db | 146  | GlnAlaAlaGlnThrArgAlaAlaValGlnThrPheAlaHisAlaLeuSerLeuSerIle    | 165  |
| OY | 715  | GGTTTTCACCTGACACACACAGCGGAGCTTATCTAAGGCTATTGACACAGAGAACAG       | 774  |
| Db | 166  | AspPheHisGlnSerLysPArgThrGlySerLeuSerArgValIleAspArgGlyAlaArg   | 185  |
| OY | 775  | GGTATACAGTTTGTGCTCGAGTGGCTTGGTATTAATCTTCCATCATGTTTGAAGT         | 834  |
| Db | 186  | SerMetAspPheLeuLeuArgGlyLeuValPheAsnLeuAlaProThrGlyIleGlnLeu    | 205  |
| OY | 835  | ATGCTGTGCAGTGTGTTGTTGTTATTAACAAATGGCGGCCACAGTTTGCCTTGGTAACTCT   | 894  |
| Db | 206  | IleLeuAlaAlaValValLeuAlaLysAlaTyrAspTyrArgPheAlaAlaValAlaLeu    | 225  |
| OY | 895  | GGAACACTTGGTACATACACACAGTTCACAGTTGCACGTGCACACGCTGGAGAACTGATTT   | 954  |
| Db | 226  | ValThrValAlaIleTyrGlyTyrValThrPheAlaIleSerAspTyrPheArgIleGlyHis | 245  |
| OY | 955  | AGAAATGAAGAAGCAAGCAAGATATGATGACAGTATCTGCTATAGCTACCTGCTG         | 1014 |
| Db | 246  | ArgArgAlaLeuAsnGlnAlaAspAlaGlnAlaAlaGlyArgAlaValAlaAspAlaLeuLeu | 265  |
| OY | 1015 | AATTATGAACGTGAGATATTTAAATATGAAGAATATGAACACAGATATGATGTGA         | 1074 |
| Db | 266  | AsnTyrGlnThrValLysSerPheGlyGlyGlnAlaArgAlaValAlaSerTyrGlnArg    | 285  |
| OY | 1075 | TTTTTGAAGAGTATGAGACTGCTCTCAATTAAGAAAGTACCTCTGCTGCTATGCTGAAC     | 1134 |
| Db | 286  | AlaLeuAspThrTyrGlyArgAlaAsnIleLysAlaThrGlnSerLeuAsnLeuLeuAsn    | 305  |
| OY | 1135 | TTTGGTCAAAGTGCATTTTTCAGTGTGGTTTAAACGCTATATAGTGTCTCCGCACTCAG     | 1194 |
| Db | 306  | LeuValGlnSerGlyValMetSerValGlyLeuGlyValMetAlaValLeuAlaGlySer    | 325  |
| OY | 1195 | GGAAATGGGACGATACCCTTACTGTTGAGATCTGTATGTGAATGTGAATGGACCTGTTT     | 1254 |
| Db | 326  | GlnAlaAlaHisGlyArgMetGlyProGlyAspValThrAlaAlaValLeuLeuLeu       | 345  |
| OY | 1255 | CAGCTTTCATTACCCCTGAACTTTCTGGGAACGTATATAGAGACGTACAGACACTC        | 1314 |
| Db | 346  | AsnLeuTyrAlaProLeuAsnIleLeuGlyPheAlaTyrArgGlnIleIleArgHisLeu    | 365  |
| OY | 1315 | ATAAGTATGAACACCTTGTTTACTCTACAGGTGACACCCAAATTAAAGCAAAGTG         | 1374 |
| Db | 366  | IleAspMetGlnAlaMetLeuAspLeuArgArg-----GlnAlaAlaAspValAla        | 382  |
| OY | 1375 | ATGGCATCTCCCTTGAGATCACACCAACACAGCT-----ACCTGGCC                 | 1419 |
| Db | 383  | AspAlaProAspAlaGlnAspLeuProCysAlaAspGlnArgGlyGlyValAlaValAla    | 402  |

[illegible]





Db 794 GlyArgIleValIgluThrGlyThrHisgluIleuLeuIleValArgAspGlyGlyArgTyr 813  
 Qy 2104 TCAGAAATGCGCATACACAGAGC 2127  
 Db 814 LysLysMetTrpPheGlnGlnAla 821

## RESULT 13

TI8376  
 multidrug resistance protein 2 - malaria parasite (Plasmodium falciparum)  
 C/Species: Plasmodium falciparum  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
 C/Accession: TI8376  
 R/Rubio, U.P.; Cowman, A.F.  
 Exp. Parasitol. 79, 137-147, 1994  
 A/Title: Plasmodium falciparum: the pfmdr2 protein is not overexpressed in chloroquine-r  
 A/Reference number: Z18924; MUID:9433528; PMID:7914495  
 A/Accession: TI8376  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1025 <RUB>  
 A/Cross-references: EMBL:U04640; NID:G439853; PID:G439854; PIDN:AAA21513.1  
 C/Genetics:  
 A/Gene: mdr2

## Alignment Scores:

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 5,85e-59 | Length:       | 1025 |
| Score:                 | 923.50   | Matches:      | 213  |
| Percent Similarity:    | 54.97%   | Conservative: | 141  |
| Best Local Similarity: | 33.07%   | Mismatches:   | 247  |
| Query Match:           | 21.71%   | Indels:       | 43   |
| DB:                    | 2        | Gaps:         | 13   |

AF133659 (1-2345) x TI8376 (1-1025)

Qy 364 AAAGCAAGCTTTCTTATGTGTGGCCCAAGAGCCGAT----- 405  
 Db 338 LysIleuLeuProIyLeuItrProserLysArgIleAsnMetLysGlyAsnSer 357  
 Qy 406 ---CTACAGCTAGAGTTCGCTGCGGATTTTGGGTGGTGCMAAGCCATGAT 462  
 Db 358 IleuAsnGlnTrpIleValIleuIlePheLeuPheIleValSerIleValPheSer 377  
 Qy 463 ATTGTGTTCCCTTCATGTTTAAATATGCTGTAGACAGCTCAACAGATGCGGAAC 522  
 Db 378 ValIleSerProIleIyLeuGlyIleValIleSerAsnGlnValIleuLysSer----- 395  
 Qy 523 ATGTGCAACCTGATGATGACCAATACAGTTCACACCATGCAACAGCATTTGATT 582  
 Db 396 -----LeuSerSerSerValIyTrpIleuGlyLeuIyValIlePhePhePheIle 412  
 Qy 583 GCGTATGATGATCAAGAGCTGAGCTGCTTTTAAAGAGTTGCAATGAGTATT 642  
 Db 413 -----SerLysPheLeuLysGlnValIleCysGlyValIleuPhe 424  
 Qy 643 GCGAAGTAGCCAGAAATTCATCCGAAGATAGCAAAAATGCTTTCATCTTAC 702  
 Db 425 SerGlnValGlnIleuSerAlaPheIleGluIleuGlnIleuSerIlePheGlnThrPheHis 444  
 Qy 703 AACCTGATCTGGGTTTTCACCTGACAGACAGAGGAGCTTATCTTACGCTATTGAC 762  
 Db 445 AsnLeuSerTrpGluTrpIySerSerLysAsnSerGlyIleMetArgIleValAsp 464  
 Qy 763 AGAGAAACAAGGGATATGTTGTCTGAGTCTTGTGATTTAATCTTCCATCCATC 822  
 Db 465 ArgGlnTrpGlnSerAlaAsnAsnLeuMetSerValIleuMetTrpIleIlePheAla 484  
 Qy 823 ATGTTTGA-----GTGATGCTTGTGAGTGTGTTGTATTAACAATGCGGTGCCAG 876  
 Db 485 ThrIleGlnIleuIleThrCysIleIlePheIlePheLysTrpLysAsnSerLeu 504  
 Qy 877 TTTCGTTTGGAACCTTGGAAACCTTGATACACAGCATTCACAGTTGACAGTCA 936  
 Db 505 GlySerValIleuPheIleGly---LeuThrLeuTrpIleIySerThrIleLysIleThr 523

Qy 937 CGGTGAGAACTAGATTAGTAATGAATGAACAAGCAGATATGATGAGATATGCT 996  
 Db 524 LysTrpArgLysLysIleArgThrIleAlaAsnIleuSerAsnValIyTrpHisAspIle 543  
 Qy 997 GCTATAGACTGACTGCTGAATTTAAGAACTGTAAGTATTTAATGAAGATATGA 1056  
 Db 544 AlaHisAspSerLeuThrAsnIyGlnAsnValIyTrpPheSerAsnGlnLysPheGln 563  
 Qy 1057 GCACAGAGATATGATGATTTTGAACAGATATGAGCTGCTCATTTGAAGAAATGACCT 1116  
 Db 564 IleLysLysPheCysAsnAlaLeuSerAsnTrpHisArgTrpAsnLysIleLeuAsn 583  
 Qy 1117 ACTGTGCTATGCTGAACCTTTGCTCA-----AGTGCATTTTCAAGTGC 1161  
 Db 584 SerLeuGlyIleLeuAsnThrValGlnGlnPheIleuAsnGlyThrLeuPhePheThr 603  
 Qy 1162 GGTTAACAGCTATATATGCTGCTGCCAGTCAGGAAATTTGCGAGGTACCTTACTGT 1221  
 Db 604 LeuLeuCysValIleIyTrpMetIleValIleGluGlySerAspProGlyThrPheIle--- 622  
 Qy 1222 GGAATCTAGTAATGGAATGAGCTCTTTTCAGCTTCATACCCCTGAACCTTCTG 1281  
 Db 623 ---SerValValValIyTrpTrpSerAsnValPhe-----AlaProLeuSerIleLeu 638  
 Qy 1282 GGAATCTATATAGAGAGACTAGACAGACTCATATATGAACACTTTACTCTA 1341  
 Db 639 GlyThrLeuTrpAlaThrIleIleLysSerPheThrAspIleSerAspIleAspIle 658  
 Qy 1342 CTCAGGTAGACACCCAAATTAAGACMAAGATGAGCATCTCCCTTCAGATACACA 1401  
 Db 659 LeuArgAspLysIleAspIleSerAsnAspLysAsnLeuLysAsnPheAspLeuThrSer 678  
 Qy 1402 CAGACA-----GCTACCGTGCCTTGTATATGCTATTTGTAATGATTGAG 1449  
 Db 679 GlnGlnLysLysPheGlyValSerIleGlnPheAsnAsnValHisPheAsnTrpProThr 698  
 Qy 1450 GGC-----CAGAAAGCTCTTAGTGATATCCCTTGAATCCCTGAGGAAAGAAATG 1503  
 Db 699 GlnProLeuHisThrSerLeuLysAspIleAsnIleIyTrpIleLysProGlyThrCys 718  
 Qy 1504 GCCATTTAGAGAGTATGCTGCTGACGGAAGAAAGACAAATAGTACGCTATTATTCGCTTC 1563  
 Db 719 AlaLeuValIyHisThrGlySerGlyLysThrThrIleSerLeuLeuTrpArgPhe 728  
 Qy 1564 TATGAGCTCAAAAGGTATGATTTATCTGCTGCTGCTCAAAATATACAAATGACGCTG 1623  
 Db 739 TyrAsp---SerLysGlyGlnIleLysIleGlyLysArgAsnIleAsnGlnTrpThrArg 757  
 Qy 1624 GAAAGCTTCGAGAGGCGAGGAGGAGTACCTCAGAGATGCTGCTCCATTAATACT 1683  
 Db 758 AsnSerIleArgAsnIleIleGlyIleValIleProGlnAspThrIleuPheAsnGlnSer 777  
 Qy 1684 ATTATTTACAACCTCTTATATGAAACATCAGTCTTCACTGAGAGATGTATGACGTG 1743  
 Db 778 IleLysTrpAsnIleLeuTrpGlyLysLeuAspAlaThrGlnIleuIleGlnAla 797  
 Qy 1744 GCAAAATTAAGCTGACTTATGATGCAATTTCTGAAATGCCCAATGATGACCCAA 1803  
 Db 798 ValLysSerIleGlnIleuTrpAspPheIleGlnSerLeuProLysTrpAspThrLeu 817  
 Qy 1804 GTAGGGGACGAGAGCAAGCTTTCAGAGAGGAGAAAGCAAGAGATGACATTTGCAGA 1863  
 Db 818 ValGlyAspLysGlyValLysLeuSerGlyGlnArgGlnArgIleSerIleAlaArg 837  
 Qy 1864 GCCATTTTGAAGACCCCGCATATCTATATGATGAAGCTATCTCATGTTAGATTGC 1923  
 Db 838 CysLeuLeuLysAspProLysIleValIlePheAspGlnAlaThrSerSerLeuAspSer 857  
 Qy 1924 ATTAATGAAGACTATTTCTTGTGCTCATGAAGATGCTCAAAACAGAACTTCTATT 1983  
 Db 858 ArgThrGlnTrpLeuPheGlnIyAlaValGlnAspLeuArgLysAsnArgThrIleIle 877



QY 1768 GCAATTCCTCGAATGCCATGATATGACACCAAGTAGGGGAACGAGACTCAAGCTT 1827  
 Db |||||  
 Db 639 LyslethrSerleupProgluIlytrAlathrMetvalGlygluArgGlyLeuLysleu 658  
 QY 1828 TCAGAGAGAGAAAGCAAGAGTAGCAATTCAGAGAGCATTTTGAAGGACCCCGCATC 1887  
 Db |||||  
 Db 659 SerGlygluIyuglmglnArgvalAlalIleAlaIthrIleLeuLysLysProGlnPhe 678  
 QY 1888 ATACTATGATGAAAGCTACTTCACTCGTTAGATTGATTCGTAAGAGACTATTCTTGGT 1947  
 Db |||||  
 Db 679 IlePheLeuArgIuAlathrSerAlaLeuMetPheProthrgluArgAlaIleGlnLys 698  
 QY 1948 GCCATGAAGATGTGTGTCACACAGAACTCTATTTCATTGACACAGATGTCAACA 2007  
 Db |||||  
 Db 699 CysLeuIyuglLysCysLysSerArgThrIyAlvalAlaIleArgLeuSerThr 718  
 QY 2008 GTGGTGTATGATGAGATGAATATGTCTTGATGATGAGTAAAGTAAAGCCAGCGTAAAC 2067  
 Db |||||  
 Db 719 ValValAsnAlaAspLeuIleLeuValLeuAspLysGlyIleIleLeuGluArgGlyAsn 738  
 QY 2068 CACCATGGTTTCTTGTCTTAACCTCATATGATCTATTCAAGATGTGACATACAGAGC 2127  
 Db |||||  
 Db 739 HistLysGluLeuLeuAlaGln---GlnGlyThrIytrAlaSerMetIyPgluAlaGlnIle 757  
 QY 2128 AGCCGTGTGCAAGAACATGATTAACCCCAATGGAGACCAAGAAAGAAATATATCCAAA 2187  
 Db |||||  
 Db 758 Ala-----GluGlnArgAlaLysSerIleGluLeu 767  
 QY 2188 GAGAGGAA 2196  
 Db |||||  
 Db 768 GlyGluGln 770  
 RESULT 15  
 F84172  
 ABC transport protein [imported] - Halobacterium sp. NRC-1  
 C/Species: Halobacterium sp. NRC-1  
 C/Date: 02-Feb-2001 #sequence.Revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C/Accession: F84172  
 R/Ny, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Laeky, S.  
 ; Leitbauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablbc  
 Jung, K.H.; Alam, M.; Freilae, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A/Authors: Hou, S.; Daniels, C.J.; Demais, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A/Title: Genome sequence of Halobacterium species NRC-1.  
 A/Reference number: A84160; MID:20504483; PMID:11016950  
 A/Accession: F84172  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-642 <STO>  
 A/Cross-references: GB:AE004437; NID:G10579770; PIDN:AA018746.1; GSPDB:GNO0138  
 C/Genetics:  
 A/Gene: trp1  
 Alignment Scores:  
 Pred. No.: 1.35e-44 Length: 642  
 Score: 725.00 Matches: 205  
 Percent Similarity: 48.15% Conservative: 120  
 Best Local Similarity: 30.37% Mismatches: 250  
 Query Match: 17.04% Indels: 100  
 Db: 2 Gaps: 17  
 AF133659 (1-2345) x F84172 (1-642)  
 QY 319 GACCCA---AAAGAGGTTAAAGATGTGATCTCGGAAATCATAAAGCATGCTT 375  
 Db |||||  
 Db 12 AspProhegluIuglnAlaArgAlaAspValAlaAspAsnAlaMetval-----GlnLeu 28  
 QY 376 TCTATGTGTGCCCAAGACAGGCCAGATCTACAGCTGAGTGTGCGATTCGCTGGGA 435  
 Db |||||  
 Db 29 PheAspIytrGlyArgAspHis---SerPheGlnAlaValAlaValAlaValAlaSer 47  
 QY 436 TTTTGGGTGTGCAAGGCCATGAATATTTGTGTTCCCTTCATGTTAAATATGCTGTA 495  
 Db |||||

Db 48 ValPhe-----AlaArgValLeuAspLeuAlaProProValLeuLeuGlyLeuAlaIle 65  
 QY 496 GAGAGCCTCAACAGATGTGCGAAAATGCTGTAACCTGATGATGACCAATATACAGTT 555  
 Db |||||  
 Db 66 AspSerValIleGln-----GlyAsnLysAlaPheLeuProheLeuProGlnSerVal 83  
 QY 556 -----GCAACATGGCAAGCAGTT 576  
 Db |||||  
 Db 84 ValProSerSerLysProAspArgLeuLeuPheMetGlyGlyLeuIleAlaGlySerPhe 103  
 QY 577 CTGATGGCTATAGGTGATATCAAGAGCTGAGCTGTTTATTAACGAAGTCCGAATATGCA 636  
 Db |||||  
 Db 104 LeuLeuGly-----AlaAlaIleHisIytrPheArgAsnIytr 115  
 QY 637 GTATTGGCAAGTAGCCAGATTCATGCAAGATATACCAAAATATGCTTCTTCAT 696  
 Db |||||  
 Db 116 GlyPheAsnSerPheSerGlnHisIleGlnHisArgValArgThrAspThrIyAspLys 135  
 QY 697 CTTCAACAACCTGATCTGGTTTCACTGAGCAGACAGACGGGA-----GCT 744  
 Db |||||  
 Db 136 MetGlnArgLeuAsnMetAspPhePheAlaThrLysGlnThrGlyGluMetSerIle 155  
 QY 745 TTAATCAAGCTATTGACACAGAGAAACAGGGTATC-----AGTTT 786  
 Db |||||  
 Db 156 LeuSerAsnAspValAsnArgLeuGluArgPheLeuAsnAspIyLeuAsnSerAlaPhe 175  
 QY 787 GTCCTAGTCTTGGATTTAATCTTCTCCATCATGTTGAAGTATGATGCTTGCAGT 846  
 Db |||||  
 Db 176 ArgLeuSerValMetVal-----LeuAlaIle 184  
 QY 847 GGTGTTTGTATTACAATGCGGTGCCAGTTGCTTGTGTAACCTTGGAACACTTGGT 906  
 Db |||||  
 Db 185 GlyValIytrLeuPheValAlaAsnIytrGlnLeuAlaValLeuThrMetLeuProValPro 204  
 QY 907 ACATACACAGATTCACAGTGTGACACAGCTGAGAGACTGATTTAGAAATGAATG 966  
 Db |||||  
 Db 205 IleIleAlaLeuPheThr-----TyrArgPheValAsnAlaIleGlnPro 219  
 QY 967 AACAAAGCATATGATGATGAGGATGATGCTGATGACTGACTGCTG----- 1014  
 Db |||||  
 Db 220 LysIytrAlaAspValArgSerSerValGlyHisLeuAsnSerArgLeuGluAsnAsnLeu 239  
 QY 1015 ---AATTATGAACCTGATATTTTAATGAAGATATGAAACACAGATATGAT 1071  
 Db |||||  
 Db 240 GlyGlyIleGlnValIleLysThrSerAsnThrGluArgIyGluSerAspArgValAsp 259  
 QY 1072 GCAATTTTGAACGATGAGACTGCTTCAATGAAGATCACTGACTGCTATGCTG 1131  
 Db |||||  
 Db 260 AspValSerGlnIytrPheAspAla----- 268  
 QY 1132 AACTTGTGCAAGTCTATTTTCACTGCTGCTTAAACGTTATATGTCGCCAGT 1191  
 Db |||||  
 Db 269 AsnIytrGlyAlaIleThrIleArgIleLysPhePheProAlaLeuArgIleIleSerGly 288  
 QY 1192 CAGGGAATTGT----- 1203  
 Db |||||  
 Db 289 ValGlyPheValLeuThrPheValIleGlyGlyIleIytrValAlaIleThrGlyProPheMet 308  
 QY 1204 -----GCAAGTACCTTACTGTTGAGATCTAGTAATGAGTGAATGACTGCTTTGAG 1257  
 Db |||||  
 Db 309 PhePheSerGlyThrLeuAspProGlyGluPheValIytrPheIleLeuLeuSerGlnGln 328  
 QY 1258 CTTTCAATACCTGTAACCTTCTGGAAGCTGATATATAGAGACTGACAGACACTCAT 1317  
 Db |||||  
 Db 329 PheIleIytrProMetAlaGlnPheGlyGlnIleIleAsnMetIytrGlnArgAlaArgAla 348  
 QY 1318 GATATGACACCTTGTATTCTACTCAAGGTAGACACCAATTAAGACAAGGTGATG 1377  
 Db |||||  
 Db 349 SerSerGluArgIlePheGlyLeuMetAsnGluProSerArgIleGluGluAsnProAsp 368  
 QY 1378 GCATCTCCCTTCAGATCACACACAGACAGACTCCGTCCTTGTATATGTCATTTT 1437  
 Db |||||  
 Db 369 AlaAspIleLeuValVal-----AspArgIyGlyValValIytrAspAspValArgPhe 386



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2003, 05:55:00 ; Search time 17.5 Seconds

(without alignments)  
6847.099 Million cell updates/sec

Title: AF133659  
Perfect score: 4254  
Sequence: 1 ATGGCCCTGCTCGCATGCA.....TATTAAAAAATCATACATT 2345

Scoring table:  
BLSUM62 Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 140259 seqs, 2554876 residues

Total number of hits satisfying chosen parameters: 280518

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.spool/AF133659/runat\_10022003\_155442\_29603/app\_query.fasta.1.2503  
-DB=published.Applications\_AA -QMT=fastan -SUFIX=rapb -MINMATCH=0.1  
-LOOPCH=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=AF133659 @CGN 1.1 10 @runat\_10022003\_155442\_29603  
-NCPU=6 -ICPU=3 -NO XLPXY -NO MMAP -LARGQUERY -NEG SCORES=0 -WALT -LONGLOG  
-DEV TIMEOUT=120 -WALT TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database: 1: /cgn2\_6/prodata/2/pubppaa/US08 NEW PUB. pep.\*

- 2: /cgn2\_6/prodata/2/pubppaa/PCT\_NEW PUB. pep.\*
- 3: /cgn2\_6/prodata/2/pubppaa/US06 NEW PUB. pep.\*
- 4: /cgn2\_6/prodata/2/pubppaa/US06 PUBCOMB. pep.\*
- 5: /cgn2\_6/prodata/2/pubppaa/US07 NEW PUB. pep.\*
- 6: /cgn2\_6/prodata/2/pubppaa/US07 PUBCOMB. pep.\*
- 7: /cgn2\_6/prodata/2/pubppaa/PCTUS PUBCOMB. pep.\*
- 8: /cgn2\_6/prodata/2/pubppaa/US08 PUBCOMB. pep.\*
- 9: /cgn2\_6/prodata/2/pubppaa/US09 NEW PUB. pep.\*
- 10: /cgn2\_6/prodata/2/pubppaa/US09 PUBCOMB. pep.\*
- 11: /cgn2\_6/prodata/2/pubppaa/US10 NEW PUB. pep.\*
- 12: /cgn2\_6/prodata/2/pubppaa/US10 PUBCOMB. pep.\*
- 13: /cgn2\_6/prodata/2/pubppaa/US60 NEW PUB. pep.\*
- 14: /cgn2\_6/prodata/2/pubppaa/US60 PUBCOMB. pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description      |
|------------|--------|-------------|--------|-------|------------------|
| 1          | 1247   | 29.3        | 836    | 10    | US-09-953-688A-7 |
| 2          | 1201.5 | 28.2        | 475    | 9     | US-10-076-157-4  |
| 3          | 1183   | 27.8        | 574    | 10    | US-09-953-688A-1 |
| 4          | 736.5  | 17.3        | 640    | 9     | US-09-976-059-9  |

|    |       |      |      |    |                     |                    |
|----|-------|------|------|----|---------------------|--------------------|
| 5  | 699.5 | 16.4 | 578  | 10 | US-09-815-242-12501 | Sequence 12501, A  |
| 6  | 685   | 16.1 | 621  | 9  | US-09-738-626-6020  | Sequence 6020, Ap  |
| 7  | 685   | 16.1 | 621  | 10 | US-09-953-253-2     | Sequence 2, Appli  |
| 8  | 654   | 15.4 | 582  | 10 | US-09-815-242-14006 | Sequence 14006, A  |
| 9  | 651   | 15.3 | 582  | 10 | US-09-815-242-10115 | Sequence 10115, A  |
| 10 | 632   | 14.9 | 587  | 9  | US-10-260-877-26    | Sequence 26, Appl  |
| 11 | 632   | 14.9 | 587  | 10 | US-09-815-242-10960 | Sequence 10960, A  |
| 12 | 618   | 14.5 | 603  | 10 | US-09-815-242-5202  | Sequence 5202, Ap  |
| 13 | 615.5 | 14.5 | 1275 | 10 | US-09-749-340-6     | Sequence 6, Appli  |
| 14 | 593.5 | 14.0 | 1247 | 9  | US-09-738-626-4751  | Sequence 4751, Ap  |
| 15 | 593   | 13.9 | 1195 | 10 | US-09-873-409-6     | Sequence 6, Appli  |
| 16 | 590.5 | 13.9 | 571  | 10 | US-09-815-242-10619 | Sequence 10619, A  |
| 17 | 590   | 13.9 | 760  | 10 | US-09-833-017-26    | Sequence 26, Appl  |
| 18 | 585   | 13.8 | 1276 | 10 | US-09-866-866A-6    | Sequence 6, Appli  |
| 19 | 582.5 | 13.7 | 766  | 9  | US-10-072-621-6     | Sequence 6, Appli  |
| 20 | 582.5 | 13.7 | 766  | 10 | US-09-795-693-17    | Sequence 17, Appl  |
| 21 | 578.5 | 13.6 | 1280 | 10 | US-09-866-866A-4    | Sequence 4, Appli  |
| 22 | 575.5 | 13.5 | 1394 | 9  | US-10-101-388-3     | Sequence 3, Appli  |
| 23 | 573.5 | 13.5 | 1276 | 10 | US-09-866-866A-8    | Sequence 8, Appli  |
| 24 | 572.5 | 13.5 | 1280 | 9  | US-10-072-621-7     | Sequence 7, Appli  |
| 25 | 572.5 | 13.5 | 1280 | 10 | US-09-866-866A-2    | Sequence 2, Appli  |
| 26 | 569.5 | 13.4 | 1222 | 10 | US-09-873-409-5     | Sequence 5, Appli  |
| 27 | 567   | 13.3 | 1272 | 10 | US-09-769-097-4     | Sequence 4, Appli  |
| 28 | 566   | 13.3 | 1272 | 10 | US-09-769-097-2     | Sequence 2, Appli  |
| 29 | 565.5 | 13.3 | 583  | 10 | US-09-815-242-13458 | Sequence 13458, A  |
| 30 | 562.5 | 13.2 | 1334 | 9  | US-09-758-828-2     | Sequence 2, Appli  |
| 31 | 562   | 13.2 | 1280 | 9  | US-10-044-671-2     | Sequence 2, Appli  |
| 32 | 554.5 | 13.0 | 656  | 9  | US-09-738-626-4577  | Sequence 4577, Ap  |
| 33 | 535.5 | 12.6 | 659  | 10 | US-09-873-409-1     | Sequence 1, Appli  |
| 34 | 535.5 | 12.6 | 812  | 10 | US-09-873-409-2     | Sequence 2, Appli  |
| 35 | 535.5 | 12.6 | 1058 | 10 | US-09-873-409-4     | Sequence 4, Appli  |
| 36 | 528.5 | 12.4 | 748  | 9  | US-09-870-759-41    | Sequence 41, Appli |
| 37 | 513   | 12.1 | 400  | 10 | US-09-765-272-190   | Sequence 190, App  |
| 38 | 509   | 12.0 | 1263 | 9  | US-09-882-694-11    | Sequence 11, Appl  |
| 39 | 485   | 11.4 | 577  | 9  | US-09-738-626-4578  | Sequence 4578, Ap  |
| 40 | 472.5 | 11.1 | 514  | 10 | US-09-873-409-8     | Sequence 8, Appli  |
| 41 | 470.5 | 11.1 | 646  | 10 | US-09-841-132-567   | Sequence 567, App  |
| 42 | 464.5 | 10.9 | 659  | 10 | US-09-841-132-497   | Sequence 497, App  |
| 43 | 461.5 | 10.8 | 551  | 10 | US-09-815-242-11394 | Sequence 11394, A  |
| 44 | 449   | 10.6 | 541  | 10 | US-09-873-409-7     | Sequence 7, Appli  |
| 45 | 447.5 | 10.5 | 339  | 9  | US-09-764-884-25    | Sequence 25, Appl  |

ALIGNMENTS

RESULT 1  
US-09-953-688A-7  
; Sequence 7, Application US/09953688A  
; Patent No. US20020102649A1  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Yue, Henry  
; APPLICANT: Reddy, Koopa  
; APPLICANT: Gorgone, Gina  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Azimzai, Yalda  
; APPLICANT: Patterson, Chandra  
; APPLICANT: Baughn, Mariah R.  
; TITLE OF INVENTION: HUMAN TRANSPORT PROTEIN HOMOLOGS  
; FILE REFERENCE: PF-0555 US  
; CURRENT APPLICATION NUMBER: US/09/953,688A  
; CURRENT FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: 09/113,427  
; PRIOR FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 836  
; TYPE: PRT  
; ORGANISM: RATTUS NORVEGICUS  
; FEATURE:  
; OTHER INFORMATION: 2962567, Genbank  
US-09-953-688A-7

## Alignment Scores:

|                        |           |               |     |
|------------------------|-----------|---------------|-----|
| Pred. No.:             | 7,54e-104 | Length:       | 836 |
| Score:                 | 1247.00   | Matches:      | 280 |
| Percent Similarity:    | 54.94%    | Conservative: | 126 |
| Best Local Similarity: | 37.89%    | Mismatches:   | 247 |
| Query Match:           | 29.31%    | Indels:       | 86  |
|                        |           | Gaps:         | 13  |

AF133659 (1-2345) x US-09-953-688A-7 (1-836)

```

Oy 28 CGCTGGGCGGCGCGCGCTCTTTCGAAAAGCCGCGCACTCCGCGATTCTGATC--- 84
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 133 ArgGlnSerLeuAlaMetGlyValTTrpMetLysPheArgHisSerLeuGlyLeuLeuLeu 152
Oy 85 -----CGGCTTACGCTCTGTAAAGCGGCTCAGGTCCGAG 120
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 153 LeuTrpThrValThrPheAlaIleGluAsnLeuValLeuValSerTrpAsnSerProGln 172
Oy 121 TGGAGGCCCATCACTCGCGCGCTTGGGAACCGCTCGAGCTACAGATTCAGATCA 180
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 173 Trp----- 173
Oy 181 TTAATAAGTATCATGCGAGATTTGGAAAAGCAATTCAGACAGTTCTTGAATGCT 240
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 174 -----TrpTrpSerArgAlaAspLeuGlyGlnGlnValGlnPhe----- 186
Oy 241 GCAAAAGGCTCTCCAGGTATGGCCAGTATAGAAAAGAGACA----- 282
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 187 -----GlyLeuTrpValLeuArgTyMetThrSerGlyGlyLeuPheIleLeu 202
Oy 283 ---TGTTGG----- 288
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 203 GlyLeuTrpAlaProGlyLeuArgProGlnSerTyThrLeuHisValAsnGluGluAsp 222
Oy 289 ---CATGCTATGCGAGAGAGAGACTCCACAGACCCAAAGAAAGGTTAAAGATGTT 345
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 223 GlnAspGlyGlyArgAsnGlnGlyArgSerThrAspProArgSerThrTrpArgAspLeu 242
Oy 346 GATTACTCGGAAAATCATAAAGCAATGCTTTCTATGTTGTGGCCCAAAACAGCGCCAGAT 405
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 243 GlyArgGlyAsp-----LeuArgLeuLeuSerGlyTyLeuTrpProArgGlySerProSer 260
Oy 406 CTACAGAGCTAGAGTGGCATTTTCGCGGAGATTTTGGGTGGTGAAGGCGCATGAATAT 465
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 261 LeuGlnLeuThrValLeuLeuGlyMetGlyLeuMetGlyLeuAspArgAlaLeuAsnVal 280
Oy 466 GTGTTCCCTTCATGTTAAATATGCTGTAGACAGCTTCACACAGATGTCCGAAAAACATG 525
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 281 LeuValProIlePheTyArgAspIleValAsnLeuLeuThr----- 294
Oy 526 CTGAACCTGAGTGAAGCAACA-----AATACAGTTGCACCAATGAGCAACAGATTTCTG 579
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 295 -----SerLysAlaIleProTrpSerSerLeuAlaTrpThrValThrTyThrValPhe 311
Oy 580 ATTAGCTAT-----GATGATCAAGAGCTGAGAGCTGCTTTTAAACAGATTTGGAAT 633
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 312 LeuLysPheLeuGlnGlyGlyGlyThrGlySerThrGlyPheValSerAsnLeuArgThr 331
Oy 634 GCAGTATTGGCAAGTAGCCCAAGATTCATCCAGAAATAGCCAAATAATGCTTTCTC 693
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 332 PheLeuTrpIleArgValGlnGlnPheThrSerArgGlyValGlnLeuArgLeuPheSer 351
Oy 644 CATCTTCACAACTGATGTTGGTTTTCCTCGAGCAGACAGACGGAGCTTATCTAG 753
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 352 HisLeuHisGluLeuSerLeuArgTrpHisLeuGlyArgArgThrGlyGlyValLeuArg 371
Oy 754 GCTATTGACAGAGAAAGGAGTATCAGTTTGTCTGAGGCTTGTGATATTATATCT 813
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 372 IleValAspArgGlyThrSerSerValThrGlyLeuLeuSerTyLeuValPheAsnIle 391
Oy 814 CTTCCCATCATGTTGAAGTGAATGCTTGTGAGTGTTGTTGATTAATATATG----- 867
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

Db 392 IleProThrLeuAlaAspIleIleIle-----GlyIleIleTyTrpSerMetPhePhe 409
Oy 868 GATGCCAGATTGCTTGGTGTGAACCTTGAAACCTTGATACATGACAGACTTACAGATT 927
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 410 AsnAlaTrpPheGlyLeuIleIleValPheLeuGlyMetSerLeuTyLeuIleThrIle 429
Oy 928 GCAGTCACACGGTGAGAGAACTGATTTAGAAATAGAAATGAAACAAAGACAGATATGATGCA 987
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 430 MetValThrGluTrpArgGlyAlaLysPheArgArgAspMetLeuThrGlnIleAsnAlaThr 449
Oy 988 GGTATGCTGCTATAGACTACTGCTGATATGAAATCTGGAAGTATTTAAATAAAGAA 1047
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 450 ArgAlaArgAlaValAspSerLeuLeuAsnPheGlnThrValLysTyTrpAsnAlaGlu 469
Oy 1048 AGATATGAGACACAGAGATATGATGATTTTGAAGACGTATGAGACTGCTTATGAAA 1107
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 470 GlyTyArgLeuGluArgTyArgGluAlaIleLeuLysPheGlnGlyLeuGluTrpLys 489
Oy 1108 AGTACTTACTCTGCTATGCTGATGCTGAACTTTGGTCAAAAGTCTATTTTCAAGTCCGTTTA 1167
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 490 SerThrAlaSerLeuValLeuLeuAsnGlnThrGlnAsnMetValIleGlyPheGlyLeu 509
Oy 1168 ACAGCTATATAGTGCTGCTGCCAGTCAAGGAATTTGGCAGGTACCCTTACTGTGAGAT 1227
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 510 LeuAlaGlySerLeuLeuCysAlaTyArgPheValSerGluArgArgLeuGlnValGlyAsp 529
Oy 1228 CTAGTATGATGAGATGAGACTGCTTTTTCAGCTTTCATTACCCCTGAACTTTCTGGGAAT 1287
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 530 PheValIleuPheGlyThrTyThrIleThrGlnLeuTyMetProLeuAsnTrpPheGlyThr 549
Oy 1288 GTATATAGAGAGACTAGACAGACCTCATAGATATGACACCTTGTATTACTTCACTCAAG 1347
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 550 TyrTyArgMetIleGlnThrAsnPheIleAspMetGluAsnMetPheAspLeuLeuLys 569
Oy 1348 GTAGACACCCAAATTAAGACAAAGATGAGTGCATCTCCCTTCAGATCAGACACACAGACA 1407
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 570 GluGluThrGluValLysAspValProGlyValIleTyProLeuArgPhe-----HisLys 587
Oy 1408 GCTACCGGCGCTTGTGATATGTCATTTGATATACATTTGAGGCGCAGAAAGCTTATAGT 1467
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 588 GlyArgValGluPheGluAsnValHisPheSerTyArgAlaAspGlyArgGluThrLeuGln 607
Oy 1468 GGAATATCTTTGGAAGTCCCTGACAGAAAGAAAGTGGCCATTTGAGAGATGAGTGTCA 1527
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 608 AspAlaSerPheThrValMetProGlyGlnThrValAlaLeuValGlyProSerGlyVala 627
Oy 1528 GCGAAAGCAACAATAGTAGAGCTATTTATTTGCTTCTATGAGCTTCAAAAGGTAGCATT 1587
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 628 GlyLysSerThrIleLeuArgLeuLeuPheArgPheTyArgPheIleSerSerGlyCysIle 647
Oy 1588 TATCTGCTGTCAAAATATATCAAGATGTGAGCTGGAAAGCTTTCGAGGGCAGGTGGA 1647
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 648 ArgIleAspGlyGlnAspIleSerGlnValThrGlnIleSerLeuArgSerHisIleGly 667
Oy 1648 GTGTACTCTCAGATGCTGCTCTTCCATTAATCTATTTATTAACAACCTTATATGGA 1707
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 668 ValValProGlnAspThrValLeuPheAsnAspThrIleAlaAsnIleArgTyGly 687
Oy 1708 AACATCAGTCTTCACTCGAGGAAGTGTATGCAAGTGCAGAAATTAAGCTGACCTTCAAT 1767
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 688 ArgValThrAlaGlyAspSerGlnIleGlnAlaIleAlaGlnAlaIleGlyIleHisAsp 707
Oy 1768 GCAATTTCTGGAATGCCACATGATGATGACACCCAAAGTGGGAGAGAGCTCAAGCTT 1827
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 708 AlaIleLeuSerPheProGluGlyTyGluThrGlnValGlyGluArgGlyLeuLysLeu 727
Oy 1828 TCAGAGAGAGAAAAGCAAGAGTAGCAATTCAGAGCAATTTTGAAGACCCCCAGCT 1887
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 728 SerGlyGlyGluLysGlnArgValAlaIleAlaArgThrIleLeuLysAlaProAspIle 747
Oy 1888 ATACTATATGATGAGCTACTTCACTCGTATGATTCATTAAGAGACATTAATCTTGGT 1947
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 748 IleLeuLeuAspGluAlaThrSerAlaLeuAspThrSerArgGlnArgAlaIleGlnAla 767

```



```

QY 1948 GCCATGAAGATGTCGTCACAGAACTTCTATTTTCATTCGACACAGATTGTCAACA 2007
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 768 SerleuAlaValCysThrAsnArgThrPheIleValAlaHisArgLeuSerThr 787

QY 2008 GTGGTGTGATGAGTGAATCATGTCTGTGATCAGGGTAAAGCCGAAACGNGTACC 2067
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 788 ValAlaSerAlaAspGlnIleLeuValIleuYsaSpGlyCysIleIleGluArgGlyArg 807

QY 2068 CACCATGGTTGCTGCTTAACCTCATAGTATCTATTCAGAAATGTGGCATACACAG 2124
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 808 HisGluAlaLeuLeuSerArg---GlyGlyValTyrAlaGluMetIrrPginLeuGln 825

RESULT 2
US-10-076-157-4
; Sequence 4, Application US/10076157
; Publication No. US20030027309A1
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Suelberger, Harald
; APPLICANT: Joeffken, Hans Wolfgang
; APPLICANT: Doval, Jose Luis Revuelta
; APPLICANT: Jimenez, Alberto;
; TITLE OF INVENTION: Genes of purine biosynthesis from Ashbya Gossypii and the use the
; TITLE OF INVENTION: 1n
; FILE REFERENCE: 48684DIV
; CURRENT APPLICATION NUMBER: US/10/076,157
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 09/212,247
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: WordPerfect v. 6.1
; SEQ ID NO 4
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Ashbya gossypii
; US-10-076-157-4

Alignment Scores:
Pred. No.: 7 37e-100 Length: 475
Score: 1201.50 Matches: 234
Percent Similarity: 71.74% Conservative: 96
Best Local Similarity: 50.87% Mismatches: 127
Query Match: 28.24% Indels: 3
DB: 9 Gaps: 2

AF133659 (1-2345) x US-10-076-157-4 (1-475)

QY 757 ATTGACAGAGAACAGAGGGATGATCTTTGTCTGAGTGTCTTGATTATCTTCTT 816
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MetcAspArgGlyCysArgGlyIleSerTyrValLeuSerAlaMetValPheHisIleIle 20

QY 817 CCATCATGTTTGAAGTGAAGCTGTGTCAGTGTGTTTGTATTAACAATGCGGTCCAG 876
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 ProIleThrPheGluIleSerMetValCysGlyIleLeuThrTyrGlnPheGlyAlaSer 40

QY 877 TTTGCTTTGGAACCTTGAACACTTGTGATACATACACAGATTCACAGTTGCAGTACA 936
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 41 PheIleAlaIleThrPheSerThrMetLeuLeuTyrSerIlePheThrPheArgThrThr 60

QY 937 CGGTGAGAACTAGATTAGATGAATGAACAAAGCAGATTAATGACGAGTAATGCT 996
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 AlaTyrArgThrArgPheArgArgAspAlaAsnGlyAlaAspAsnGlyAlaAlaSerVal 80

QY 997 GCTATGACTCACTGCTGAATATGAACACTGTGAAGTATTTTAATGAAGAAATATGAA 1056
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 81 AlaLeuAspSerLeuIleAsnPheGluAlaValTyrTyrPheAsnAsnGlnTyrIleu 100

QY 1057 GCACAGAGATATGATGATTTTGAACAGATATGACCTCTTCATTTGAAGAAAGTACTCT 1116
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 101 AlaAspTyrThrHisThrSerLeuMetCysTyrArgAspSerGlnIleuValSerGln 120

```

```

QY 1117 ACTGTGGCTATGCTGAACCTTGTGCAAGTGTCTATTTTCAGTGTGCTTGAACGATA 1176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 SerLeuAlaPheLeuSerThrGlyGlnAsnLeuIlePheThrAlaLeuThrAlaMet 140

QY 1177 ATGTGCTGCCAGTCAAGGAATTTGGCAGGTACCTTTCCTGTGGAGATCTAGTAAG 1236
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 MetTyrMetAlaCysAsnGlyValMetGlnGlySerLeuThrValGlyAspLeuValLeu 160

QY 1237 GTAATGAGACTGCTTTTTCAGTTTCATTTACCCCTGAACTTCTGTGGAATCTATAGA 1296
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 IleAsnGlnLeuValPheGlnLeuSerValProLeuAsnPheLeuGlnSerValTyrArg 180

QY 1297 GAGACTAGACAAACACTCATATGATGAACACCTGTGTTACTCAAGTAGACACC 1356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 AspLeuGlnSerLeuIleAspMetGluSerLeuPheLeuGlnIlySaAsnGlnVal 200

QY 1357 CAATTTAAACAAAGATGATGATCTCCCTTCAATACACACACAGACGTACCGTG 1416
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 ThrIleYsaSerProAsnAlaGlnAsnLeuProIle---HisYsaProLeuAspIle 219

QY 1417 GCCTTGATATGATGATTTGAAATACATTGAGGCGCAGAAAGTCTTAGTGAATATCC 1476
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 ArgPheGlnAsnValThrPheGlyTyrAspProGluArgArgIleLeuAsnAsnValSer 239

QY 1477 TTGGAAGTCCCTGACAGAAAGAAAGTGGCCATTGTAGAGTGGTTCAGGAAAGC 1536
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 PheThrIleProAlaGlyMetCysThrAlaIleValGlyProSerGlySerGlyYsaSer 259

QY 1537 ACAATAGTGAAGCTATATTTCCTTCTATAGACCTTCAAAAGGTGACATTATCTTGT 1596
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 ThrIleLeuYsaLeuValPheArgPheTyrGluProGlnGlnIlyArgIleLeuValGly 279

QY 1597 GGTCAAAATATACAGATGAGCTGGAAGCTTGGAGGCGAGTGGAGTGGTACT 1656
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 GlyThrAspIleArgAspLeuAspLeuSerLeuArgGlyAlaIleGlyValValPro 299

QY 1657 CAGATGCTGTCTCTTCAATATATCTATTATTAACCTCTTATATGGAACATCAGT 1716
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 GlnAspThrProLeuPheAsnAspThrIleTyrGlnAsnValYsaPheGlyAsnIleSer 319

QY 1717 GCTTCACTGAGGAATGATGATGACGTGGCAAAATTTGCTGAGCTTATGATGCAATCTT 1776
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 SerSerAspAspIleIleLeuArgAlaIleGlyAlaGlnLeuThrYsaLeuGln 339

QY 1777 CGAATGCCATGATATGACACCCAAAGTAGGGAAGGAGACTCAAGCTTTGAGAGGA 1836
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 AsnLeuProIlyGlyAlaSerThrValValGlyArgGlyLeuMetIleSerGlyGly 359

QY 1837 GAAACCAAGATAGCAATTTGCAAGAGCCATTTTGAAGAGCCCCAGCATACTCTAT 1896
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 GlnYsaGlnArgLeuAlaIleAlaArgValLeuLeuYsaPheAlaProLeuMetPhePhe 379

QY 1897 GATGAAGCTATTCATGTTAGTTGATTAAGTGAAGACTATTTCTGTGTCATGAAG 1956
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 AspGluAlaThrSerAlaLeuAspThrHisThrGlnGlnAlaLeuLeuHisThrIleGln 399

QY 1957 GAT-----GTGTCAAACAGAACTTCTATTTTATTTGACACAGATTTGTAACAGG 2010
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 GlnAsnPheSerSerAsnSerYsaThrSerValTyrValAlaHisArgLeuArgThrIle 419

QY 2011 GTTGATGAGATGATGATCTTGTGATCAGGGTGAAGGAGCCAGCCAGTATCCAC 2070
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 AlaAspAlaAspYsaIleIleValLeuGlnGlnGlySerValArgGlnGlnGlyThrHis 439

QY 2071 CATGTTTCTGCTGTAACCTCATATGATATTCAGAAATGTGCATACACAGACGACC 2130
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 440 SerSerLeuLeuAlaSerGlnGlySerLeuTyrArgGlyLeuThrAspIleGlnGlnAsn 459

RESULT 3
US-09-953-688A-1
; Sequence 1, Application US/09953688A
; Patent No. US20020102649A1

```

```

: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Yue, Henry
: APPLICANT: Reddy, Roopa
: APPLICANT: Gorgone, Gina
: APPLICANT: Corley, Neil C.
: APPLICANT: Azimzai, Yalda
: APPLICANT: Patterson, Chandra
: APPLICANT: Baughn, Mariah R.
: TITLE OF INVENTION: HUMAN TRANSPORT PROTEIN HOMOLOGS
: FILE REFERENCE: PR-0555 US
: CURRENT APPLICATION NUMBER: US/09/953, 688A
: PRIOR FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: 09/113,427
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 574
: TYPE: PRT
: ORGANISM: HOMO SAPIENS
: FEATURE:
: OTHER INFORMATION: 2074412, ISLNOT01
: US-09-953-688A-1

Alignment Scores:
Pred. No.: 3,68e-98 Length: 574
Score: 1183.00 Matches: 250
Percent Similarity: 62.12% Conservative: 114
Best Local Similarity: 42.66% Mismatches: 194
Query Match: 27.81% Gaps: 28
DB: 10

AF133659 (1-2345) x US-09-953-688A-1 (1-574)

QY 430 CTGGATTTTGGGTGGTGCAAGGCCATGATATTGGTCCCTTCATGTTAAATAT 489
DB 1 MetGlyLeuMetGlyValArgArgAlaLeuAsnValPheValProIlePheTyrArg--- 19
QY 490 GCTGTAGACAGCTCAACCAAGATGTCCGGAACATGCTGCAACCTG-----AGTATGCA 543
DB 20 -----AsnIleValAsnLeuLeuThrGluLysAla 29
QY 544 CCA---AATACAGTTCGAACCATGCAACAGCA-----GTTCTGATTGGC 585
DB 30 ProThrPheSerLeuAlaIrrPheValThrSerTyrValPheLeuLysPheLeuGlnGly 49
QY 586 TATGGTGTATCAAGAGCTGAGCTGCTCTTTTAAAGATTGCAAGTCAATGATTGGC 645
DB 50 GlyGlyThrGlySerThrGly-----PheValSerAsnLeuArgThrPheLeuTrpIle 67
QY 646 AAGTAGCCCAAGATTCATCCGAAGATAGCCAAAATGTTCTTCTCATCTTCACAAAC 705
DB 68 ArgValGlnGlnPheThrSerArgArgValGluLeuLeuIlePheSerHisLeuHisGlu 87
QY 706 CTGGATCTGGGTTTCACTGAGCAGACAGAGGGAGCTTATCTTAAGCTATTGACAGA 765
DB 88 LeuSerLeuArgIrrPheSerLeuGlyArgGlnGlyGluValLeuArgIleAlaAspArg 107
QY 766 GGAACAAGGGATATCATGTTTGTCTGAGTGTCTGTTGTTAATCTTCCCATCATG 825
DB 108 GlyThrSerSerValThrGlyLeuLeuSerTyrLeuValPheAsnValIleProThrLeu 127
QY 826 TTGGAAGTATCTGTGTCAGTGTGTTTGTATTCACAAATGC-----GTCGCCAGTTT 879
DB 128 AlaAspIleIleIle-----GlyIleIleTyrPheSerMetPhePheAsnIleIrrPhe 145
QY 880 GCTTGTGTAACCTTGGAACACTGGTATCATACAGATTCAAGTTCAGTTCAGTACAGG 939
DB 146 GlyLeuIleValIlePheLeuLysMetSerLeuTyrLeuThrIleValValThrGlu 165
QY 940 TGGAGAACTAGATTAGATAGAAAAGCAAGACAGATATGATGACAGTATGCTGCT 999
DB 1166 TrrArgThrLysPheArgAlaMetAsnThrGlnLysAsnAlaThrArgAlaArgAla 185
QY 1000 ATAGACTCAGCTGCAATTAAGAACTGTGAAGATTATTAATAAGATATGAGCA 1059
DB 186 ValAspSerLeuLeuAsnGlnThrValLysTyrTyrAsnAlaGluSerTyrGluVal 205
QY 1060 CAGAGATATGATGATTTTGAAGACGTATGAGCTGCTTCATTGAAAGTCACTTACT 1119
DB 206 GluArgTyrArgGluAlaIleIleLysTyrGlnGlnLysLeuGlnTrrPheSerSerAlaSer 225
QY 1120 CTGGCTATGCTGAACCTTTGGTCAAGTGTATTTTCACTGTCGGTTTAAACGCTATATG 1179
DB 226 LeuValLeuLeuAsnGlnThrGlnAsnLeuValIleGlyLeuGlyLeuAlaGlySer 245
QY 1180 GTGCTCGCAGCTGAGGAGATGTGGCAGGTACCTTACTGTGTGAGATCTAGTATGATG 1239
DB 246 LeuLeuLysAlaTyrPheValThrGlnGlnLysLeuGlnValGlyAspTyrValLeuPhe 265
QY 1240 AATGACTGCTTTTTCAGCTTTTCATTACCCCTGAACTTTCTGGGAAGCTATATAGAG 1299
DB 266 GlyThrTyrIleIleGlnLeuTyrMetProLeuAsnThrPheGlyThrTyrTyrArgMet 285
QY 1300 ACTAGACAGACATCATATGATATGACACACTTGTACTTACTTCAAGTAGACACCAA 1359
DB 286 IleGlnThrAsnPheIleAspMetGluAsnMetPheAspLeuLeuLysGluGlnThrGlu 305
QY 1360 ATTAAAGCAAGAGATGGCATCTCCCTTCAGATCAACACAGACAGACAGTACCGTGC 1419
DB 306 ValLysAspLeuProGlyAlaGlyProLeuArgPhe-----GlnLysGlyArgIleGlu 323
QY 1420 TTTGATATGTGCAATTTTGAATATACATTAGAGGCGCCAGAAAGCTTATGTAATCTTT 1479
DB 324 PheGluAsnValHisPheSerTyrAlaAspGlyArgGluThrLeuGlnAspValSerPhe 343
QY 1480 GAATCCCTGCAAGAAAGAAAGTGGCCATTGTAGAGTAGTGGTCAAGGAAAGACACA 1539
DB 344 ThrValMetProGlyGlnThrLeuAlaLeuValGlyProSerGlyAlaGlyLysSerThr 363
QY 1540 ATAGAGAGCTATATTTCGCTTATGAGCTCAAAAGGTACATTTATCTTGCTGT 1599
DB 364 IleLeuArgLeuLeuPheArgPheTyrAspIleSerSerGlyCysIleArgIleAspGly 383
QY 1600 CAAATATACAGATGTGAGCTGGAAGCTTGGAGGCGGAGGAGTGTATCTCAG 1659
DB 384 GlnAspIleSerGlnValThrGlnAlaSerLeuArgSerHisIleGlyValValProGln 403
QY 1660 GATGCTGCTCTTCATATATCTATTTATTCACACTCTTATATGGAACATAGTGTCT 1719
DB 404 AspThrValLeuPheAsnAspThrIleAlaAspAsnIleArgTyrGlyArgValThrAla 423
QY 1720 TCACCTGAGAGGTATGAGCTGAGCAAAATTGACTGAGCTTCATGATGCAATCTCGA 1779
DB 424 GlyAsnAspGluValGluAlaAlaIleGlnAlaIleGlyIleHisAspAlaIleMetAla 443
QY 1780 ATGCCACATGATATGACACCAAGTAGGGGAACGAGACTCAAGCTTTCAGAGAGAA 1839
DB 444 PheProGluGlyTyrArgThrGlnValGlyGluArgGlyLeuLysSerGlyGlyGlu 463
QY 1840 AAGCAAGAGTAGCAATTCAGAGCACTTTGAAGACCCCGCCAGTCATATCTATGAT 1899
DB 464 LysGlnArgValAlaIleAlaIleArgThrIleLeuLysAlaProGlyIleIleLeuLeuAsp 483
QY 1900 GAAGCTACTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1959
DB 484 GluAlaThrSerAlaLeuAspThrSerAsnGlnArgAlaIleGlnAlaSerLeuAlaLys 503
QY 1960 GTGTCAAAACAGACACTTATTTTATTCACACAGATTTGCAAGCTGTGATGCA 2019
DB 504 ValCysAlaAsnArgThrThrIleValValAlaHisArgLeuSerThrValValAsnAla 523
QY 2020 GATGAATATCTGCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2079
DB 524 AspGlnIleLeuValIleLysAspGlyCysIleValGluArgGlyArgHisGluAlaLeu 543

```





```
Db 306 GlnSerPheAlaSerMetAspArgValPheGlnLeuIleAspGluAspTyrAspIleLeu 325
1366 GACAAAGATGAGTGCCTCCCTCAGATCACACACAGACGATACCGTCCCTTGGAT 1425
Db 326 AsnGlyValGlyAlaGlnProIleGluLeu-----LysGlnGlyValGlyIleAsp 343
1426 AATGTGATTTGAAATACATTAGAGGCGCAG--AAAGCTTGAATGAAATACCTTTGAA 1482
Db 344 HisValSerPheGlnTyrAsnAspAsnGluAlaProIleLeuLysAspIleAsnLeuSer 363
1483 GTCCCTGACGAAAGAAAGTGGCCATTGTAGAGGTATGGGTCAAGGAAAGCAATA 1542
Db 364 IleGluLysGlyGluThrValAlaPheValGlyMetSerGlyGlyLysSerThrLeu 383
1543 GTGAGGCTATTTGCTTATGAGCTCCAAAGAGCTACATTTATTCCTGCTGCTCA 1602
Db 384 IleAsnLeuIleProGlyPheTyrAspValThrSerGlyGlnIleLeuIleAspGlyHis 403
1603 AATATACAGATGTGAGCTCGAAAGCCTTCGAGGCGACGTGGAGTGGTACCTCAGAT 1662
Db 404 AsnIleLysAspPheLeuThrGlySerLeuArgAsnGlnIleGlyLeuValGlnGlnAsp 423
1663 GCTGCTCTTCCATATATATCTATTATTACAACTCTTATATGAAACATGATGCTTCA 1722
Db 424 AsnIleLeuPheSerAspThrValLysGluAsnIleLeuGlyArgProThrAlaThr 443
1723 CCTGAGAGATGTATGACGTGGCAAAATATCTGATCTTCATGATGATCTTCGATG 1782
Db 444 AspGluGluValValGluAlaAlaLysMetAlaAsnAlaHisAspPheIleMetAsnLeu 463
1783 CCACATGATATGACACCCACAGTAGGAGACGACATCAAGCTTTGAGAGAGAAAG 1842
Db 464 ProGlnGlyTyrAspThrGluValGlyLysValGlyValLysLeuSerGlyGlyGlnLys 483
1843 CAAAGAGTACCAATTTGCAAGGCCATTTTGAAGACCCCGCAGTCATATCTTATGATGA 1902
Db 484 GlnArgLeuSerIleAlaArgIlePheLeuAsnAsnProProIleLeuIleLeuAspLys 503
1903 GCTACTTCATGCTTATGATGATTAAGTACGAAAGACTATCTTGGTCCCATGAAGAGCTG 1962
Db 504 AlaThrSerIleAlaLeuAspLeuGluSerIleIleGlnGluAlaLeuAspValLeu 523
1963 GTCAACACAGAACTTCTATTTTATTCATGACACAGATTTGCAAGTGGTATGACAGAT 2022
Db 524 SerLysAspArgThrThrLeuIleValAlaHisArgLeuSerThrIleThrHisAlaAsp 543
2023 GAAATCATTTGTTGGATCAGGGTAAAGGTAGCCGAAACGTGTACCCACATGTTTGCTT 2082
Db 544 LysIleValValIleGluAsnGlyHisIleValGluThrGlyThrHisArgGluLeuIle 563
2083 GCTAACCTCATAGTATTTATTCAGAAATGTGCAATACACAGAC 2127
Db 564 AlaLys---GlnGlyAlaTyrGluHisLeuTyrSerIleGlnAsn 577

RESULT 6
US-09-738-626-6020
; Sequence 6020, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OR INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
```

```
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6020
; LENGTH: 621
; TYPE: PRN
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6020

Alignment Scores:
Pred. No.: 2,35e-53 Length: 621
Score: 685.00 Matches: 193
Percent Similarity: 50.23% Conservative: 129
Best Local Similarity: 30.11% Mismatches: 247
Query Match: 16.10% Indels: 72
DB: Gaps: 19

AF133659 (1-2345) x US-09-738-626-6020 (1-621)
Qy 355 AAATCATAAAGCAATGCTTTCTTATGTGTGGCCCAAGACAGGCGCATCTACAGACT 414
Db 6 ArgIleLeuLysThr---ThrSerAlaLeuThrProTyrTyrLeuGlyIleIleValVal 24
415 AGAGTTCGCAATTCGCTGGGATTTTGGGTGGTGCAGAAAGCCATGAATATTGGTCCC 474
Db 25 SerIleValIleAla-----AlaLeuSerLeuLeuSerPro 36
475 TTGATGTTAAATATGCTGTACAGAGCTC--AACCAAGTTCGGAAACATCTGTAAC 531
Db 37 PheIleLeuArgGluAlaThrAspSerIleValSerAlaValThrGlySer-----Asn 54
532 CTGAGTATGCA---CCAATACAGTTGCAACCACTGACAGACAGACAGATTCGATGCTAT 588
Db 55 ThrValAspAlaValAlaThrArgThrIleIlePheLeuAlaLeuAlaLeuPheVal----- 72
589 GGTGTATCAAGAGCTGAGCTGCTTTTATTAACCAAGTTGCAATGAGTA----- 639
Db 73 -----AlaSerPheLeuAsnThrValMetThrAsnIleGlyTyr 86
640 TTGGCAGAGTACCCGAGATTCATCCGAAATA--GCCAAATATGCTTTCTCAT 696
Db 87 IleGlyAspValMetAlaSerArgMetArgGlnIleLeuAlaThrArgTyrTyrAlaLys 106
697 CTTCAACAACCTGATCGGGTTCACCTGAGCAGACAGACGGGAGCTTATCTAAGGCT 756
Db 107 LeuLeuAlaLeuProGlnLysTyrPheAspAsnGlnValThrGlyThrIleIleAlaArg 126
757 ATTGACAGAGAAACAGAGGATATCAGTTTGTCTGTAGCTTGTGATTAATCTTCTT 816
Db 127 LeuAspArgSerIleAsnGlyIleThrGlnPheMetGlnSerPheSerAsnAspPhe 146
817 CCATC-----ATGTTGAAGTATCTGTCTGCTGTTTGTATTACAAATGC 867
Db 147 PrometLeuIleThrMetValAlaValLeuIleIleSerAlaIlePheTyrTyrProLeu 166
868 GGTGCCAGTTGCTTTGATCCCTTGAACACTGTGTACATACACAGATTCACAGATT 927
Db 167 Ala-----IleLeuAlaMetLeuPheProIleTyrMetTyrPheThr 181
928 GCAATCACA-----CGGTGAGAACTAGATTGATAGAAATGAACAAAGCAGATTAAT 981
Db 182 AlaLeuThrSerLysArgTyrGln---LysTyrGluGlyGlyLysAsnHisGluIleAsp 200
982 GATGCAAGTATGCTGTATAGACTACCTGCAATTATGAAACTGTGAAGTATTTTAAT 1041
Db 201 ValAlaAsnGlyArgPheAlaGluValValGlyGlnValLysValLysSerPheVal 220
```

```
Oy 1042 AATGAAGATATGAAGCAGAGATATGATGA---TTTGAAGACGTATGAGACTGCT 1098
    |||
Db 221 AlagIuThrArgGluLeuValAspPheGlyArgTyGlyLeuValAlaIleThr 240
Oy 1099 TCATTGAAAAGT-----ACCTCTACTCTGGCTATGCTGAACTTTGGTCAA 1143
    :|||
Db 241 ArgProGlnSerGlyTrpTrpHisArgMetAspThrLeu-----Arg 254
Oy 1144 AGTGCATATTTGTCGCTTAAACAGCTATATGTCCTGCCAGTCAGGAATTGTC 1203
    |||
Db 255 GlYAlaIleAsnLeuMetIlePheLeuAlaIleIleLeuIlePheTyArgThrLeu 274
Oy 1204 GCAGGTACCTTACTGTTGGAATCTAGTAATGTGAATGACTGCTTTTCACTTCA 1263
    |||
Db 275 HisGlyHisPheThrIleGlyAspMetValMetLeuIleGluValThrMetAlaGln 294
Oy 1264 TTAACCCCTGAATTTCTGGAACTGTATATAGAGACTAGACAGCACTCATAGATAG 1323
    |||
Db 295 GlnProValTyMetMetSerTyxIleValAspSerAlaGluAlaIleAlaGlySer 314
Oy 1324 AACACCTTGTACTCTACTC-----AAGTAGACACCCAAATTAA 1365
    |||
Db 315 ArgAspTyxPheGluValMetAlaGlnValGluProThrAlaAsnLysGluLeuVal 334
Oy 1366 GACAAAGTATGACATCT-----CCCTTGATGATCACA 1398
    |||
Db 335 AspAlaThrLeuAlaSerAspThrProArgIleSerValGlyThrProAlaIleLeuPro 354
Oy 1399 CCACAGACAGCTACCGCTGCTTTGATATATGTCATTTTGAATACATTAGAGCCAGAAA 1458
    |||
Db 355 AlaGlyGluProAlaMetGluPheLysAsnValThrPheAlaTyxGluIleLysPro 374
Oy 1459 GTCTTAAGTGAATATCCCTTGAAGTCCCTGCAGAGAAAGAGTGGCTTAGAGGT 1518
    |||
Db 375 ValIleSerAspValSerIleThrAlaArgHisGlyGluArgIleAlaLeuValGlyGlu 394
Oy 1519 AGTGGTCAAGGAAAAACAAATAGTAGAGCTATTATTTCGCTTATGAGCTCAAAAG 1578
    |||
Db 395 SerGlyGlyGlyLysSerThrLeuValAlaLeuLeuGlyLeuTyxLysProAsnSer 414
Oy 1579 GGTACATTTATCTGCTGCTGCAAAATATACAGATGTAGAGCTGGAAGCTTGGAGG 1638
    |||
Db 415 GlySerLeuAlaValCysGlyValAspValLysAspLeuThrSerGluLeuLeuArgAla 434
Oy 1639 GCAGTGGAGTGGTACCTCAGATGCTGCTCTTCATATATCTTATTATACACCTC 1698
    |||
Db 435 SerValGlyValValPheGlnAspAlaSerLeuPheSerGlySerIleAlaGluAsnIle 454
Oy 1699 TTATATGAACATCATAGTCTTACCTGAGAGAGTGTATGACGTGGCAAAATTAAGCTGGA 1758
    |||
Db 455 AlaTyxGlyArgProGlyValaThrArgGluGluIleIleGluValAlaIleLysAlaAsn 474
Oy 1759 CTTTATGATGCAATTTCTGGAATGCCACATGATATGACACCCAGTAGGGAGAGAGA 1818
    |||
Db 475 AlaHisGluPheIleSerAlaPheProGluGlyTyxGluThrValValGlyLysArgGly 494
Oy 1819 CTCAGAGCTTTGAGAGAGAGAAAGAGAGATAGACATGAGAGCCATTTGGAAGAC 1878
    |||
Db 495 LeuLysLeuSerGlyGlyGlnLysGlnArgValSerValAlaArgAlaMetLeuLysAsp 514
Oy 1879 CCCCAGTCATACTCTATGATGAAGCTTATCATCGTTAGATTGATTAAGAGACT 1938
    |||
Db 515 AlaProLeuLeuValLeuAspGluAlaThrSerAlaLeuAspThrLysSerGluAla 534
Oy 1939 ATTCTTGGTGCATGAAGATGTGTCACACAGAACTTCTATTCTTCTGACACAGA 1998
    |||
Db 535 ValGlnAlaGlyLeuGluIleLeuMetGluAsnArgThrThrLeuMetIleAlaHisArg 554
Oy 1999 TTGTCAACCTGATGATGATGATGAATATCATGTCTTGTGATCAGGATGAAGCCGGA 2058
    |||
Db 555 LeuSerThrIleAlaGlyValAlaSerThrIleValIleThrIleGlnAsnGlyArgValGluGlu 574
Oy 2059 CGTGTATCCACCATGTTGCTTGTCTTAAACCTCATATGATCTATTACAGAAATGTGGCAT 2118
```

```
Db 575 ValGlySerProThrGlu---LeuAlaValSerGlyGlyIleTyxSerGluLeuLysArg 593
Oy 2119 AACAGAGACCGCTGTGCAGACACATGATTAACCCAAATGAGAGCAAAAGAAAT 2178
    |||
Db 594 LeuThrAsnSerThrAla-----GluAlaAspArgGluArg 605
Oy 2179 ATA 2181
    |||
Db 606 Leu 606

RESULT 7
US-09-953-259-2
; Sequence 2, Application US/09953259
; Patent No. US20020115159A1
; GENERAL INFORMATION:
; APPLICANT: FARMICK, Mike
; APPLICANT: HUTTMACHER, Klaus
; APPLICANT: PFEFFERLE, Walter
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES CODING FOR THE ATR61 PROTEIN
; FILE REFERENCE: 213903USOX
; CURRENT APPLICATION NUMBER: US/09/953,259
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: DE10045579.4
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-953-259-2

Alignment Scores:
Pred. No.: 2,35e-53 Length: 621
Score: 685.00 Matches: 193
Percent Similarity: 50.23% Conservative: 129
Best Local Similarity: 30.11% Mismatches: 247
Query Match: 16.10% Indels: 72
DB: Gaps: 19

AF133659 (1-2345) x US-09-953-259-2 (1-621)
Oy 355 AAAATCATATAAGCAATCTTCTTATGTGTGCCCCAAAGACAGCCAGATCTACGAGCT 414
    |||
Db 6 ArgIleLeuLysThr---ThrSerAlaLeuThrProTyxTrpLeuGlyIleIleValVal 24
Oy 415 AGAGTGGCATTTGCTGGAGATTTTGGGTGCTGCAAGGCCATGATATTTGGTTCCC 474
    |||
Db 25 SerIleValIleAla-----AlaLeuSerLeuLeuSerPro 36
Oy 475 TTCAATGTTAATATGCTGTAGACAGCTC---AACCAAGATGGGGAAACATGCTGAGAC 531
    |||
Db 37 PheIleLeuAspGluAlaThrAspSerIleValSerAlaValIleThrGlySer-----Asn 54
Oy 532 CTGAGTGATGCA---CCAAATACAGTTGCAACATGCGAACAGCAGAGCTTGTATGCTAT 588
    |||
Db 55 ThrValAspAlaValaThrArgThrIleIlePheLeuAlaLeuAlaLeuPheVal----- 72
Oy 589 GGTGTATCAAGAGCTGAGCTGCTTTTATTAAAGAGTTGCAAAATGCAAGTA----- 639
    |||
Db 73 -----AlaSerPheLeuAsnThrValMetThrAsnIleGlyTyx 86
Oy 640 TTGGCAAGGTAGCCCAAGATTCATCCGAGAGTA---GCCAAAAATGCTTCTTCAT 696
    |||
Db 87 IleGlyAspValMetAlaSerArgMetArgGlnIleLeuAlaThrArgTyxTrpAlaLys 106
Oy 697 CTTCAACCACTGATCTGTGGTTTTCACCTGAGACAGACAGCAGGAGCTTATCTTAAGGCT 756
    |||
Db 107 LeuLeuAlaLeuProGlnLysTyxPheAspAsnGlnValThrGlyThrIleIleAlaArg 126
Oy 757 ATTGACAGAGGAACAGGGGATACGTTTGTCTGTAGTGCTTTGGATTTAAATCTTCTT 816
    |||
```

```

Db 127 LeuAspArgSerIleAsnGlyIleThrGlnPheMetGlnSerPheSerAsnAsnPhePhe 146
QY 817 CCATC-----ATGTTGAAGTGAAGCTTGTACGTGGTGTGTTGTTATTAACAATGC 867
Db 147 ProMetIleuIleThnMetValAlaValLeuIleIleSerAlaIlePheTyrTrpProIeu 166
QY 868 GGTGCCACGTTTGCTTTGGTAAACCTTGGACATTCACAGCATTCACAGCTT 927
Db 167 Ala-----IleIleuValIleMetLeuPheProIleTyrMetTrpLeuThr 181
QY 928 GCAGTCACA-----CGGTGAGAACATAGATTGAAATGAAAGAAACAGCATAT 981
Db 182 AlaLeuThrSerLysArgTrpGln--LysTyrGlnGlnLysValAsnHisGlnIleAsp 200
QY 982 GATCAGAGTAATGCTGCTATGACTACTGCTGAATTAATGAACCTGTAAGATTTAT 1041
Db 201 ValAlaAsnGlyArgPheAlaGlnValAlaGlnValLysValValLysSerPheVal 220
QY 1042 AATGAAGATATGAAGCAGACAGATATGATGA--TTTGAAGACGTATGAGACTGCT 1098
Db 221 AlaGlnThrArgGlnLeuAlaAspPheGlyArgTyrGlnLysThrValAlaIleThr 240
QY 1099 TCATTGAAAGT-----ACCTTACTCTGGCTATGCTGAACCTTGCTCA 1143
Db 241 ArgProGlnSerGlyTrpTrpHisArgMetAspThrLeu-----Arg 254
QY 1144 AGTCTATTTTCAGTGCCTGTTTAAACAGCTAATGAGTGTGCGCAGTCAAGAAATGTC 1203
Db 255 GlyAlaIleAlaLeuMetIleIlePheLeuAlaIleHisIleuLeuIlePheTyrArgThrLeu 274
QY 1204 GCAGGTACCTTACTGTTGAGATCTAATGTAATGTAATGAGTCTGTTTCACTTCA 1263
Db 275 HisGlnHisPheThrIleGlyAspMetValMetLeuIleGlnLeuValThrMetAlaGln 294
QY 1264 TTACCCCGAACCTTTCGGGAACTGTATATAGACAGCATACACAAAGCATCATGATATG 1323
Db 295 GlnProValLysMetMetSerTyrIleValAspSerAlaGlnAlaIleAlaGlySer 314
QY 1324 AACACCTGTTTACTTACTACT-----AAGTACAGACCCCAATTAA 1365
Db 315 ArgAspTyrPheGlnValMetAlaGlnGlnValGlnProThrAlaAsnHisGlnLeuVal 334
QY 1366 GACAAAGTATGCACTT-----CCCTTCAGATGACACA 1398
Db 335 AspAlaThrLeuAlaSerAspThrProArgLysSerValGlnThrProAlaIlePro 354
QY 1399 CCAAGACAGCTACCGGCGCTTGTATATGCAATTTGATATCATTTAGGCGCAGAA 1458
Db 355 AlaGlyLysProAlaMetGlnPheLysAsnValThrPheAlaTyrGlnGlnLysPro 374
QY 1459 GTCCTTAGTGAATATCTTTGAAGTCCCTCGAGAAAGAAAGGCGCATTTGTAGAGGT 1518
Db 375 ValIleSerAspValSerIleThrAlaArgHisGlnGlnAlaGlnIleAlaValGlyGln 394
QY 1519 AGTGGGTCAAGGAAACACAAATAGTGAAGCTATTATTTGCTTCTATGAGCCCAAAAG 1578
Db 395 SerGlyGlyLysSerThrLeuValAsnLeuLeuGlnLysLeuTyrLysProAsnSer 414
QY 1579 GGTAGCATTTATCTTGCTGGTCAAAATATACAGATGTGAGCCCTGGAAGACCTTGGAGG 1638
Db 415 GlySerLeuAlaValCysGlyValAspValLysAspLeuThrSerGlnGlnLeuAla 434
QY 1639 GCAGTGGAGTGAATCTGAGATGCTGCTCTTCATATATATATTTTAAACCTC 1698
Db 435 SerValGlyValValPheGlnAspAlaSerLeuPheSerGlySerIleAlaGlnAsnIle 454
QY 1699 TTATATGAAACATCACTGCTTCACTGAGGAAGTATGCAAGTGGCAAAATTAAGCTGA 1758
Db 455 AlaTyrGlyArgProGlyAlaThrArgGlnGlnIleIleGlnValAlaLysValAsn 474
QY 1759 CTTCATATGCAATTTCTTCAATGCCACATGAGATATGACACCCCAAGTACGGAAGGGA 1818
Db 475 AlaHisGlnPheIleSerAlaPheProGlnGlyTyrGlnThrValValGlyGlnArgGly 494

```

```

QY 1819 CTCAGCTTTCAGAGAGAGAAAGCAAGATGACAAATTCGACAGCCATTGTAAGGAC 1878
Db 495 LeuLysLeuSerGlyGlnLysGlnArgValSerValAlaArgAlaMetLeuLysAsp 514
QY 1879 CCCCCAGTACTCTTATGATGAAGTACTTATCATGTTAGATTGATTTACTGAAGAGACT 1938
Db 515 AlaProLeuLeuValIleuAspGlnAlaThrSerAlaLeuAspThrLysSerGlnGlnAla 534
QY 1939 ATTCTTGTCATGAAGATGTGTCAAACAGACCTTATTTTCATTTCATTCACACAGA 1998
Db 535 ValGlnAlaGlyLeuGlnLeuMetGlnAsnArgThrThrLeuMetIleAlaHisArg 554
QY 1999 TTGTCACATGCTGTTGATGACATGAATCATTTGCTGATCAGGTAAGTACCGGAA 2058
Db 555 LeuSerThrIleAlaGlyValAspThrIleValThrIleGlnAsnGlyArgValGlnGln 574
QY 2059 CGTGTACCCACCATGGTTGCTTGTCTTAACCTTCATAGTATCTATTTCAGAAATGTCAT 2118
Db 575 ValGlySerProThrGln--LeuAlaValSerGlyGlyIleTyrSerGlnLeuLysArg 593
QY 2119 ACACAGACGAGCCGTGACAGAACCATGATACCCCAAATGGAGCAAGAAAGAAAT 2178
Db 594 LeuThrAsnSerThrAla-----GlnAlaAspArgGlnArg 605
QY 2179 ATA 2181
Db 606 Leu 606

RESULT 8
US-09-815-242-14006
; Sequence 14006, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14006
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-14006

Alignment Scores:
Pred. No.: 1.4e-50 Length: 582
Score: 654.00 Matches: 187
Percent Similarity: 49.03% Conservative: 115

```



Best Local Similarity: 30.36% Mismatches: 266  
 Query Match: 15.37% Indels: 48  
 DB: 10 Gaps: 15

AF133659 (1-2345) x US-09-815-242-14006 (1-582)

```

Oy 310 CTCACACAGACCCAAAGAGGTTAAAGATGTTACTCGGAAATCATTAAGCA 369
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 MethSAsnAap-----LysAspLeuSerThrTrpGlnThrPheArgArg 15
Oy 370 ATGCTTTCTATATGTGGCCCAAGACAGGCGAGATCTACAGCTGAGCTGCCATTGC 429
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 16 Leu-----TrpProThrIleAlaProPheLysAlaGlyLeuAlaIleGly 31
Oy 430 CTGGGATTTTGGTGGTGGCAAGGCCATGATATTGGTCCCTCATGTTAAATAT 489
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 32 IleAlaLeuIleLeuAsnAlaAlaSerAspThrPheMetLeuSerLeuLeuysProleu 51
Oy 490 GCTGTAGACAGCTTCACACAGATGTCGGGAAACATGCTGAACCTGATGACCAAAAT 549
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 52 LeuAspAspGlyPheGlyLys-----ThrAspArg 61
Oy 550 ACAGTTGCAACCATGGAACACAGACAGTTCGATGGCTATGATCAAGAGCTGAGCT 609
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 62 SerValIleLeuTrpMetProLeuValIleGlyLeuMetIleLeuArgGlyIleThr 81
Oy 610 GCTTTTAAAGAA---CTTGAATGAGATATTGGCAGAGTACCAGAAATTCATC 666
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 82 SerTyrIleSerSerTyrCysIleSerTrpValSerGlyLysValValMetThrMetArg 101
Oy 667 CGAAGAAATGACCAAAATGCTTCTCCATCTTCACAACTGAGATCGGTTTCACCTG 726
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 102 ArgArgLeu-----PheGlyHisMetMetGlyMetProValAlaPhePheAsp 117
Oy 727 AGCAGACAGACGGGAGCTTATCTAAGCTATTGACAGAGCAAGAGGGGATTCAGTTT 786
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 118 LysGlnSerThrGlyThrLeu-----LeuSerArgIleThrTyrAspSerGlnGln 134
Oy 787 GTCCGAGAGCTTGGATTTAATGCTCTCCATCCCATGATTTGAAGTGAAGCTTGAGCT 846
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 135 ValAlaSerSerSerSerGlyAlaLeuIleThrValAlaArgGlnGlyAlaSerIleIle 154
Oy 847 GGT-----GTTTGTATTACAATGCGGAGCCAGCTTGGCTTGAACCTT 894
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 155 GlyLeuPheIleMetMetCHeTyrTrpSerTrpGlnLeuSerIleIleLeuValValLeu 174
Oy 895 GGAACACTTGGTACATACACAGCATTCACAGTTCACAGCAGGTGAGAACTGATTT 954
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 175 AlaProIle-----ValSerIleAlaIleArgValValSerLysArgPhe 189
Oy 955 AGAATGAAATGAACAAAGCAGATATGATGAGGT-----AATGCTGCTATAGAC 1005
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 190 ArgSerIleSerLysAsnMetGlnAsnThrMetGlyGlnValThrThrSerAlaGlnGln 209
Oy 1006 TCACGTCTGAATTAATGAACCTGGAAGTATTTAATGAAGATATGAAACACAGAGA 1065
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 210 MetLeuLysGlyHisLysGlnValLeuIlePheGlyGlnGlnValGlnThrLysArg 229
Oy 1066 TATGATGATTTTGAAGAGTATGAGAGCTCTCATTTGAAGAAAGTACTTACTGCT 1125
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 230 PheAspLysValSerAsnLysMetArgLeuGlnGlyMetLysMetValSerLysSer 249
Oy 1126 ATGCTGAACCTTGTCAAGGTCTATTTTCACTGTCGGTTTAACAGCTATATGGTGCT 1185
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 250 IleSerAspProIleIleGlnLeuIleAlaSerLeuAlaLeuAlaPheValLeuLysAla 269
Oy 1186 GCCAGTACAGGAATTTGTGACAGTACCTTACTGTTGAGTACTGATTAATGGAATGA 1245
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 270 AlaSerPheProSerValMetAspSerLeuThrAlaGlyThrIleThrValValPheSer 289
Oy 1246 CTGCTTTTACCTTTACATTACCCCTGAACTTTCTGGGAAGCTGTATAGAGAGCTAGA 1305
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 290 SerMetIleAlaLeuMetArgProLeuLysSerLeuThrAsnValAlaGlnPheGln 309
  
```

```

Oy 1306 CAAGCACTATAGATATGAACACCTTGTATTACTACTCAAGGTAGACACCCAAATTA 1365
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 310 ArgGlyMetAlaAlaCysGlnThrLeuPheAlaIleLeuAspSerGln---GlnGlyLys 328
Oy 1366 GACAAAGTATGATGATCTCCCTTCAGATTCACACCACAGACAGCTACCGTGCTTGAT 1425
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 329 AspGlu-----GlyLysArgValIleAspArgAlaThrThrLysAspLeuGluPheArg 345
Oy 1426 AATGTGATTTGAATCATGAGGCGCAAGAGTC-----CTTAGGAAATATCTTT 1479
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 346 AsnValThrPheThrTyr---ProGlyArgGlnValProAlaLeuAsnIleAsnLeu 364
Oy 1480 GAAGTCCCTGCAGAGAAAGAAAGTGGCCATTGTAGAGAGTACTGAGTCAAGGAAAGACA 1539
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 365 LysIleProAlaGlyLysThrValAlaLeuValGlyThrSerGlySerGlyLysSerThr 384
Oy 1540 ATAGTAGGCTATATTTCGCTTCTATAGAGCTCAAAAAGGTAGCATTTATCTTGCTGT 1599
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 385 IleAlaSerLeuIleThrArgPheThrAspIleAspGlnGlyHisIleLeuMetAspGly 404
Oy 1600 CAAATATACAAAGTGTAGAGCTGGAAGCCTTGGAGGCGAGTGGAGGTGCTCTCAG 1659
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 405 HisAspLeuArgGlnTyrThrLeuAlaSerLeuArgAsnGlnValAlaLeuValSerGln 424
Oy 1660 GATGCTGCTCTTCCATTAATATTATTATTAACACCTCTTATATGAAACATCAGTGCT 1719
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 425 AsnValHisLeuPheAsnAspThrValAlaAsnAsnIleAlaTyrAlaArgThrGlnGlu 444
Oy 1720 --TCACCTGAGAGAGTATGACAGTGGCAAAATTAAGTGAATTCATGATGCAATTCTT 1776
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 445 TyrSerArgGlnGlnIleGlnGlnAlaAlaArgMetAlaTyrAlaMetAspPheIleAsn 464
Oy 1777 CGAATGCCACATGATATGACCCCAAGTAGGAGAGACAGACTCAAGCTTCAGGAGGA 1836
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 465 LysMetAspAsnGlyLeuAspThrIleIleGlyGlnGlnGlyValLeuLeuSerGlyGly 484
Oy 1837 GAAAGCAAAAGATGACATTTGCAAGAGCCATTTTGAAGAGACCCCGCATCATCTTAT 1896
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 485 GlnArgGlnAlaGlyIleAlaIleAlaArgAlaLeuLeuArgAspSerProIleLeuIleLeu 504
Oy 1897 GATGAGCTACTTATCGTATGATTCGATATCTGATACAGACTATTTCTTGCCCATGAAG 1956
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 505 AspGlnAlaThrSerAlaLeuAspThrGlnAspThrArgAlaIleGlnAlaAlaLeuAsp 524
Oy 1957 GATGTGCTCAACACAGACTTCTATTTTCTGTCACACAGATTTGTAACAGTGGTGAT 2016
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 525 GlnLeuGlnLysAsnArgThrSerLeuValIleAlaHisArgLeuSerThrIleGlnGln 544
Oy 2017 GCAGATGAATCATTTGCTTGCATCAGGCTAAGGTAAGGTAGCCGAGCTGATCCCATGGT 2076
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 545 AlaAspGlnIleValAlaValGlnAspGlyIleIleValGlnGlnGlyThrHisSerGlu 564
Oy 2077 TTGCTGTAAACCTCATATGATCTATTCAGAAATGTGGCATACAG 2124
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 565 LeuLeuAlaGln---HisGlyValTyrAlaGlnLeuHisLysMetGln 579
  
```

RESULT 9  
 US-09-815-242-10115  
 ; Sequence 10115, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Karl L.  
 ; APPLICANT: Zykkind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; FILE OF INVENTION: Prokaryotes  
 ; FILE REFERENCE: EUTRA.011A

```

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ. ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 10115
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10115

Alignment Scores:
Pred. No.: 2,61e-50
Score: 651.00
Percent Similarity: 49.19%
Best Local Similarity: 30.65%
Query Match: 15,308
DB: 10
Gaps: 17

AF133659 (1-2345) x US-09-815-242-10115 (1-582)
QY 310 CTCACACAGACCCAAAGAGGTTAAAGATTTGATCTCGCAAAATCATATAAGCA 369
   |||:|||||
Db 1 MethisAsnsp-----LysAspLeuSerThrTrpGlnThrPheArgArg 15
QY 370 ATGCTTTTATGTGTGAGCCCAAGACAGCCAGAT-----CTACGAGCTAGA 417
   |||:|||||
Db 16 Leu-----TrpProThrIleAlaProPheLysAlaGlyLeuIleValAlaGly 31
QY 418 GTTCCCATTTGCGTGGATTTTGGGTGTCGCAAGGCCATGAATATGTGTTCCCTTC 477
   |||:|||||
Db 32 ValAlaIleuIleLeuAsnAlaIleSerAspThrPheMetLeuSerLeuLeuLysProLeu 51
QY 478 ATGTTTAAATATGCTGTAGACAGCCTCAACCAAGATGTCGGAAACATGCTGAACCTGACT 537
   |||:|||||
Db 52 Leu-----AspAsp-----GlyPheGly 57
QY 538 GATGCACCAATACAGTTGCAACCATGCAACAGAGCTGATGATTGGCTATGTGATCA 597
   |||:|||||
Db 58 LysThrAspArgSerValIleuValTrpMetProLeuValIleGlyLeuMetIleLeu 77
QY 598 AGAGCTGAGACTGCTTTTAAAGAA---GTTGCAATGACAGTATTGTCGCAAGGTAGCC 654
   |||:|||||
Db 78 ArgGlyIleThrSerTyrValSerSerTyrCysIleSerTrpValSerGlyLysValVal 97
QY 655 CAGAAATTCATCCGAAAGATAGCCAAATAATGTCTTTCTTCATCTTCAACACCTGATCTG 714
   |||:|||||
Db 98 MetThrMetArgArgArgLeu-----PheGlyHisMetMetGlyMetProVal 113
QY 715 GGTTTACCTGAGCAGACAGACGGAGCTTTATCTAAGGTATGACAGAGAACAGAG 774
   |||:|||||
Db 114 SerPheAspLysGlnSerThrGlyThrLeu-----LeuSerArgIleThrTyr 130
QY 775 GGATATCAGTTTGTCTGAGCTGCTTTGTTATCTTCTCCATCAGTATTGTAAGTG 834
   |||:|||||
Db 131 AspSerGlnGlnValAlaSerSerSerSerGlyAlaLeuIleThrValValArgGlnGly 150
QY 835 ATGCTTGTACAGTGT-----GTTTGTATTACAAAGCGGTGCCAGTTTGTCT 882
   |||:|||||
Db 151 AlaSerIleIleGlyLeuPheIleMetMetPheTyrTyrSerTrpGlnLeuSerIleIle 170
```

```

QY 883 TTGGTAACCTTGGAACACTTGGTAATACACAGACTTCACAGTTGACAGCGGTGG 942
   |||:|||||
Db 171 LeuIleValLeuAlaProIle-----ValSerIleAlaIleArgValVal 185
QY 943 AGAATCTAGATTTTGATAGAAATGAACCAAGCATATATGATGCAGT-----AAT 993
   |||:|||||
Db 186 SerLysArgPheArgAsnIleSerLysAsnMetGlnAsnThrMetGlyValIleThrThr 205
QY 994 GGTCTATAGACTCACGCTGAGTGAATTTGAACCTGTAAGTATTTTATATGAAGATAT 1053
   |||:|||||
Db 206 SerAlaGlnIleMetLeuLysGlyHisLysGlnValLeuIlePheGlyGlnGlnVal 125
QY 1054 GAACACAGAGATATGATGATTTTGAACCGTATAGACTGCTTCAATGAAGATACC 1113
   |||:|||||
Db 226 GlnThrLysArgPheAspLysValSerAsnArgMetArgLeuGlnIleMetLysMetVal 245
QY 1114 TCTACTGTGCTTATGCTGAACCTTGGTCAAAAGCTATTTTCACTGCTGTTAAACGCT 1173
   |||:|||||
Db 246 SerAlaSerSerIleSerAspProIleIleGlnLeuIleAlaSerLeuAlaLeuAlaPhe 265
QY 1174 ATATGGTGTGCGCAGTCAAGGAAATTTGGCAGATACCTTACTGTTGGAGATCTAGTA 1233
   |||:|||||
Db 266 ValLeuTyrAlaAlaSerPheProSerValMetAspSerLeuThrAlaGlyThrIleThr 285
QY 1234 ATGTAATGAGACTGCTTTTTCAGTTCATTCATTCACCTTGAACCTTGGGAACTGTATAT 1293
   |||:|||||
Db 286 ValValAlaPheSerSerMetIleAlaLeuMetArgProLeuLysSerLeuThrAsnValAsn 305
QY 1294 AGAGAGACTAGACAAACACTCATATGATATGAACACCTTGTTACTTACTCAAGTAGAC 1353
   |||:|||||
Db 306 AlaGlnPheGlnArgGlyMetAlaIleCysGlnThrLeuPheThrIleLeuAspSerGln 325
QY 1354 ACCCAATTTAAACAAAGATGATGCATCTCCCTTCAAGATCACACACAGACGATACC 1413
   |||:|||||
Db 326 --GlnGlnLysAspGlu-----GlyLysArgValIleGlnArgAlaThrGlyAsp 341
QY 1414 GTGCGCTTGTATATGATGCAATTTTGAATACATGAGGCGCAGAAAGTCT-----CTTAGT 1467
   |||:|||||
Db 342 ValGlnPheArgAsnValIleThrPheThrTyr---ProGlyArgAspValProAlaLeuArg 360
QY 1468 GGAATATCCTTTGAAGTCCCTGCGAGAAAGAGGCCATTTGAGAGGTAGTGAGTCA 1527
   |||:|||||
Db 361 AsnIleAsnLeuLysIleProAlaGlyLysThrValAlaLeuValGlyArgSerGlySer 380
QY 1528 GCGAAAGACATATAGTAGAGCTATATTTCGCTTTATAGACCTCAAAAAGGTAGCATT 1587
   |||:|||||
Db 381 GlyLysSerThrIleAlaSerLeuIleThrArgPheTyrAspIleAspGlnGlyGlnIle 400
QY 1588 TATCTTGCTGTCAAATATATACAAAGATGTGAGCTGGAAGCCTTGGAGGCGCAGTGGGA 1647
   |||:|||||
Db 401 LeuMetAspGlyHisAspLeuArgGlyTyrThrLeuAlaSerLeuArgAsnGlnValAla 420
QY 1648 GTGTGACTCAGAGATCTGTCTTTCATATATCTATTATTTATTAACCTCTTATATGA 1707
   |||:|||||
Db 421 LeuValSerGlnAsnValHisLeuPheAsnAspThrValAlaAsnMetIleAlaTyrAla 440
QY 1708 AACATAGTGTCT---TCACCTGAGCAAGGTATGACAGTGGCAAAATTAGCTGACCTTCAT 1764
   |||:|||||
Db 441 ArgThrGlnGlnTyrSerArgGlnGlnIleGlnGlnAlaAlaArgMetAlaTyrAlaMet 460
QY 1765 GATGCAATCTTTCGAGCAGACATGATATGACACCCAGATGAGGAGGAGCAGCTCAG 1824
   |||:|||||
Db 461 AspPheIleAsnLysMetSerPheAsnGlyLeuAspThrValIleGlyGlnSerGlyValLeu 480
QY 1825 CTTTCAGAGAGAGAAAGCAAGAGATGCAATTTGCAAGAGCCATTTTGAAGAGCCCA 1884
   |||:|||||
Db 481 LeuSerGlyGlyGlnArgGlnIleAlaIleAlaIleAlaArgAlaLeuLeuArgAspSerPro 500
QY 1885 GTCATACTCATGATAGACGTAATCTTATGCTTATGATTCGATTACTGAAAGACATATTCCT 1944
   |||:|||||
Db 501 IleLeuIleLeuAspIleAlaThrSerIleAlaLeuAspThrGlnSerGlnArgAlaIleGln 520
```

```

QY 1945 GGGTGCATGAAGGATGTGGCGAACAACAGAACTTTATTTCCTTGACACAGATTGTCA 2004
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
Db 521 AAlaAleuAspGluIleuGlnIlySerLeuValIleAlaHisArgLeuSer 540
QY 2005 ACAGTGTGATGTCACATGATAATCATCTGTCTTGATCAGGGTAAGGCCGAACTGTG 2064
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
Db 541 ThrileGlnLysAlaAspGluIleValIleValGlnAspGlyValIleValGlnArgGly 560
QY 2065 ACCCACCATGCTTGTCTGCTAACCCCTCATGATCTATTTCAGAAATGTGGCATPACAG 2124
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
Db 561 ThrHisAsnAspLeuIleuGlnHis--ArgGlyValTyrrIleGlnIleuHisLysMetGln 579

RESULT 10
US-10-260-877-26
; Sequence 26, Application US/10260877
; Publication No. US2003002181A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Chovan, Linda E.
; APPLICANT: Heselger, Paul E.
; APPLICANT: Reich, Karl A.
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
; FILE REFERENCE: 6565 US.P1
CURRENT APPLICATION NUMBER: US/10/260,877
CURRENT FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: US/09/649,145
PRIOR FILING DATE: 2000-08-25
NUMBER OF SEQ ID NOS: 137
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 587
TYPE: PRN
ORGANISM: H. influenzae
US-10-260-877-26

Alignment Scores:
Pred. No.: 1,34e-48 Length: 587
Score: 632.00 Matches: 187
Percent Similarity: 49.01% Conservative: 111
Best Local Similarity: 30.76% Mismatches: 264
Query Match: 14,86# Indels: 46
DB: 9 Gaps: 17

AF133659 (1-2345) x US-10-260-877-26 (1-587)
QY 340 GATGTTGATCTACGCGAAATCATATAAAGAATGTTCTTATGTCGCCCCAAGACAG 399
      ||| ||| ::| ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 AspPhseThrIeuGlnTrhPheLysArgLeu-----TrpProMetIleLys 25
QY 400 CCAGATCTACAGACTAGAGTTGCCATTCCCTGGGATTTTTGGGTGGTGAAGGCCATG 459
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
Db 26 Pro---PheLysAlaGlyLeuIleValSer-----GlyValAlaIleValPhe 40
QY 460 AATATTGTGCTCCC-----TTCATGTTTAAATATGCTGTAGACAGCTCTAAC 507
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
Db 41 AsnAlaIleuAlaAspSergIlyLeuIleTyrrLeuLysProLeuIleuAsp---- 58
QY 508 CAGATGTCGGGAAACATGTCGAACCTGATGATGAGCAACCAATACAGTTGACACCATGCA 567
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 -----GlyPheGlyLysAlaAsnHisSerPheLeuLysMetMet 71
QY 568 ACAGCAGTTCGATGGCTATGCTATGATCAAGAGCTGAGACTGCTTTTAAAGAGTT 627
      |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
Db 72 AlaPheValValAlaGlyMetIleIleLeuArgGlyIleTrhAsnPheIleSerAsnTy 91
QY 628 CGAATATGA--GTATTGGCAAAGTAGCCGAGAATTCATCCGAGAATATGCCAAATAAT 684
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 CysLeuAlaTrpValSergIlyLysValIleMetThrMetArgTrgArgLeu----- 108
QY 685 GTCTTTCTCATCTTTACAACTCGATCTGGGTTTTTACCCTGACAGACAGACGGAGCT 744
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

|    |      |                    |  |                    |              |                   |        |     |
|----|------|--------------------|--|--------------------|--------------|-------------------|--------|-----|
| Db | 109  | ---                | PhelyshisIeuMet                                  | Phemet             | ProValSerPhe | AerSgIlnAsnSerThr | GlyArg | 127 |
| Qy | 745  | TTATCTAAGCGTAT     | TGACACAGAAACAAGGGATACAGTTTGCTGCTGAGCGCTTGGTA     | 804                |              |                   |        |     |
| Db | 128  | Leu-----           | LeuSerArgIleThr                                  | TyrAspSerGlnMetIle | AsnSerSerSer | 144               |        |     |
| Qy | 805  | TTTATCTCTCTCCAT    | CAGTGTGAA-----GTGATCGCTTGCAAGTGAGTGTGG           | 855                |              |                   |        |     |
| Db | 145  | GlySerLeuIleThrIle | ValArgGlnGlyAlaTyrIleIleSerLeuPheAlaValMet       | 164                |              |                   |        |     |
| Qy | 856  | TATTACAAATGCGGTG   | CCCGACGTTTGCTGGTAAACCCCTTGGAACCTTGTCATACACAA     | 915                |              |                   |        |     |
| Db | 165  | PheTyrThr-----     | SerTrpGluLeuThrIleValIleuPheIleGlyProIleIle      | 182                |              |                   |        |     |
| Qy | 916  | GCATTCCACAGTTGCA   | GTCAACCGGTGAGACCTGACTTGAATAGAAATGAACAAACCA       | 975                |              |                   |        |     |
| Db | 183  | AlaValLeuIleArgLeu | ValSerLysIlePheArg--ArgLeuSerLysAsnLeuGln        | 201                |              |                   |        |     |
| Qy | 976  | GATAT-----         | GATGACGGTAAATGCGGTATACATCACTGCTGATATATGAACCTGG   | 1029               |              |                   |        |     |
| Db | 202  | AspSerMetGlyGluLeu | ThrSerAlaThrGlnMetLeuLysGlnIleLysValVal          | 221                |              |                   |        |     |
| Qy | 1030 | AAGTATTTTAAATGA    | AAAGATATGAGCACACAGATATGATGGATTTTGAAGACGTAT       | 1088               |              |                   |        |     |
| Db | 222  | LeuSerPheGlyGlnIle | ValGlnGluValIlePheAsnIleValSerAsnAspMet          | 241                |              |                   |        |     |
| Qy | 1090 | GAGACTGCTCAT       | TGAAAAGTACCTACTCTGCTGCTATGCTGTAACCTTGGTGAAGTCT   | 1148               |              |                   |        |     |
| Db | 242  | ArgArgLysSerMetLys | MetValThrAlaAsnSerIleSerAspProValGlnVal          | 261                |              |                   |        |     |
| Qy | 1150 | ATTTTCAGTGCCTG     | TTTAACAGCTATATAGTGCTGCGCAGACAGGAATTTGSCA--       | 1206               |              |                   |        |     |
| Db | 262  | IleAlaSerLeuAla    | ValAlaThrValLeuTyrLeuAlaThrThrProLeuIleAlaGlu    | 281                |              |                   |        |     |
| Qy | 1207 | GGTACCCCTACTGT     | TGGAGAGTCTAGTAATGGTGAATGAGACTGTTTTCAGCTTTCATTA   | 1266               |              |                   |        |     |
| Db | 282  | AspAsnLeuSerAla    | GlySerPheThrValValPheSerSerMetLeuAlaMetMetArg    | 301                |              |                   |        |     |
| Qy | 1267 | CCCCGAACTTTCT      | GGAAGCTGTATATAGAGACTAGACAAAGCACTCATATGATATGAC    | 1332               |              |                   |        |     |
| Db | 302  | ProLeuLysSerLeu    | ThrAlaValAlaAsnAlaGlnPheGlnSerGlyMetIleAlaCysGln | 321                |              |                   |        |     |
| Qy | 1337 | ACCTTGTTTACT       | CTACACAGGTAGACACCCAAATTAAAGCAAAAGTATGTCATCTCCC   | 1386               |              |                   |        |     |
| Db | 332  | ThrLeuPheAlaIle    | LeuAspLeuGluProGlnLysAspSP-----GlyAla            | 337                |              |                   |        |     |
| Qy | 1387 | CTTCGATCACACAC     | ACAGACAGCTACCGCTTGATTAATATGATATGATATACATT        | 1448               |              |                   |        |     |
| Db | 338  | TyrLysAlaGluPro    | AlaLysGlnLeuGlnPheLysAsnValSerPheAlaTyrGln       | 357                |              |                   |        |     |
| Qy | 1447 | --GAGGGCCAGAA      | AGTCTTACTGGAATATCTTTGAAAGTCCCTCGACGAAAGAAAGTG    | 1503               |              |                   |        |     |
| Db | 358  | GlyLysAspArgLeu    | AlaLeuAsnIleSerPheSerValProIleGlyThrVal          | 377                |              |                   |        |     |
| Qy | 1504 | GCCATTTAGAGAGT     | AATGGGTGACGGGAAAGACAAATGCTAGAGCTATATTATGGCTTC    | 1566               |              |                   |        |     |
| Db | 378  | AlaLeuValGlyArg    | SerGlySerGlyLysSerThrIleAlaAsnLeuValThrArgPhe    | 397                |              |                   |        |     |
| Qy | 1564 | TATGAGCTCAAAAG     | GTATACATTTATCTTGCTGCTCAAAATATACAAAGATGTGACCTG    | 1622               |              |                   |        |     |
| Db | 398  | TyrAspIleGluGln    | GlnGlyIleLeuLeuAspGlyValAsnIleGlnAspYrThrArgLeu  | 417                |              |                   |        |     |
| Qy | 1624 | GAAAGCTTCGAGAG     | GACGTGGAGTGTACTCAGAGATGCTGTCTCTTCCATTAATACT      | 1683               |              |                   |        |     |
| Db | 418  | SerAsnLeuArgGlu    | AsnGlyValAlaValSerGlnGlnValIleuPheAsnAspThr      | 437                |              |                   |        |     |
| Qy | 1684 | ATTATTTCACACCT     | CTTATAT--GGAAACATAGTGCCTTACCTGAGGAATGTATGCA      | 1744               |              |                   |        |     |
| Db | 438  | IleAlaAsnAsnIle    | AlaTyrAlaAlaGlnAspLysTyrSerArgGluIleIleIleAla    | 457                |              |                   |        |     |
| Qy | 1741 | GTGCGAAATTAAGT     | GTGACTTCATGATGCAATCTTCGATGCGACATGATGATGACACC     | 1800               |              |                   |        |     |
| Db | 458  | AlaAlaLysAlaAla    | TyrAlaLeuGlnPheIleGlnLysLeuProGlnValPheAspThr    | 477                |              |                   |        |     |



```
OY 1387 CTTGATCACACACAGACAGCTACCGTGCTTGTATGATGCTATTGATACATT 1446
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 338 TTYTYSALAGLPRALALAYSELGLULLEUGLUPHELVSASNAVLSERPhela1aYrGln 357
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1447 ---GAGGGCCAGAAAGTCTTAGTGAATATCTTGAAGTCCCTGCAGAAAGAAAGT 1503
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 GLYSASPGULLeuAlaleuSNANLIESerPhesValProAlAGLYSThrVal 377
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1504 GCCATTGAGAGGTAGTGGGTGAGGAAAGCAATAGTAGGCTATTATTGGCTTC 1563
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 378 AlaleuValAGLYArGserTylSerGlySerThrIleAlaSNLeuValThrArGPe 397
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1564 TATGAGCTCAAAAAGGTAGCATTTATCTTGGTCAAAATATACAAGATGAGCCTG 1623
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 398 TYRAspIleGluGlnGlyLuleuLeuSNAspGlyValASNleGlnAspYrArgLeu 417
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1624 GAAAGCTTGGAGGAGGAGTGGAGTGTACTCAAGATGCTGTCTTCATTAATACT 1683
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 SerASNLeuArgGluASNcysAlaValValSerGlnGlnValHISLeuPhesASNAspThr 437
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1684 ATTATTACAACTCTTATAT---GAAAACATCAGTGTTCACCTGAGAAAGTATGCA 1740
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 438 IleAlaASNAsnIleAlaIleAlaIleAlaAspLysTYRSerArGlnGlnIleIleAla 457
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1741 GTGGCAAAATTAGTGCATGATCATGATGCAATTTCCGAATGCCACATGATATGACACC 1800
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 AlAlaIleYsAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAla 477
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1801 CAATGAGGGAAGCAAGACTCAAGCTTTGAGAGGAGAAAGCAAGATGCAATTGCA 1860
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 478 ValIleGlyLuleuSNcylYrHISerLeuSerGlyGlyGlnArgGlnArgLeuAlIleAla 497
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1861 AGAGCCATTTTGAAGGACCCCGCTCATCTATGATGAAGCTATCTTCATCGTTAGAT 1920
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 ArGAlaLeuLeuArgASNserProValLeuIleLeuAspGlnAlaThrSerAlaLeuSNP 517
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1921 TCGATTACAGAGACTCTTCTTGGTCCATGAGTGTGTCACACAGCAACTTCT 1980
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 518 ThrGluSerGluArgAlaIleGlnSerAlaLeuGlnGluLeuLysAlaAspArgThrVal 537
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1981 ATTTTCATTGACACAGATTGTCAACAGTGTGTGATGACAGTGAATCATCTTGGAT 2040
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 538 ValValIleAlaHISArGLeuSerThrIleGluASNAlaAspGlnIleLeuValIleAsp 557
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 2041 CAGGCTAGAGTACCGGACGATGTCACCCACCATGTTGCTGTACCTCATAGTATC 2100
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 558 HISGlyGluIleArgGluArgGlyASNHisLysThrLeuLeu---GlnGlnASNGLYAla 576
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 2101 TATTCGAAATGTCGATACACAG 2124
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 577 TYRlysGlnLeuHISerMetGln 584
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-09-815-242-5202
: Sequence 5202, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: EUTRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
```

```
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 1410
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5202
: LENGTH: 603
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5202

Alignment Scores:
Pred. No.: 2,47e-47 Length: 603
Score: 618.00 Matches: 185
Percent Similarity: 49.02% Conservative: 116
Best Local Similarity: 30.13% Mismatches: 237
Query Match: 14.53% Indels: 76
DB: 10 Gaps: 18

AF133659 (1-2345) x US-09-815-242-5202 (1-603)
OY 355 AAAATCAATAAAGCAAGCTTTCTTATGTGTGGCCAAAGACAGGCGAGATCTACGAGCT 414
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 14 LysIleTYrPheArGLeuLeuGlyTYrVal-----LysProTYrIleGlyMet 29
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 415 AGAGTGGCATTTGCTGGGCAATTTTGGGTGCTGCCAAAGCCATGATATGTGCTCCC 474
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 30 PheLeuSerIleValAGlyPheIleIlePheAlaSerThrGlnProMetLeuAGly 49
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 475 TTCATGTTAAATATGCTGTAGACAGCCTCAACAGATGTGCGGAAACATGCTGAC--- 531
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 50 IleLeu---LysTYrPheValAspGlyLeuSerAsnProAspAlaIleLeuPheProAsn 68
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 532 -----CTGAGTATGACACCAATACAGTTGCAACCATGACCAAGCAGATT 576
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 ValGlnTrpOTrPheuArgAspLeuHISerValTYrAlaValProLeuIleIle 88
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 577 CTGATT-----GGCTATGCT-----GTATCAAGACT 603
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 89 LeuIleAlaIleArgGlnGlyLeuGlySerPheLeuGlyAsnPhePheLeuAlaIleVal 108
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 604 GGAGCTGCTTTTAAAGCAAGTTCGAAATGCAATATTTGSCAGATGACCCAGATTC 663
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 SerLeuGlyLeuValHISAspLeuArgValAlaLeuPheAsnLYSLeuLeu----- 125
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 664 ATCCGAAGATAGCCAAATAATGTCTTTCTCCATCTTCACAACTGAGT----- 711
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 -----ValLeuProAsnArgTYrPheAspThrHISerSerGlyHISLeuIleSer 142
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 712 ---CTGGTTTACCTGAGACAGACAGCGGAGCTTTATCTAAAGCTATTGACAGAGA 768
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 143 ArgIleThrPheAsnValThrMetValThrGlyAlaIleAlaThrAspAlaIleLYSValVal 162
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 769 ACAAG---GGTATCAGTTTGTCTGAGTCTTGGATTTATCTTCCCATCATG 825
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 163 IleArgGlyLeuLeuTh-----ValValPheLeuPheLeuTYrLeuLeu 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 826 TTGTAAGTATGCTTGCAGTGTGTTTGTATTAACAATGCGGTGCCAGTTGCTTGG 885
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 Trp-----MetAsnTrpYS----- 182
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 886 GTAACCTTGAACACTTGTATACACAGATTCACAGTTGAGTCACAGCGGTGAGCA 945
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 LeuThrLeuValMetLeuAlaIleLeuProValIleAlaValMetValThrThrAlaser 202
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

Qy 946 ACTAGATTAGAAATGAAACAA-----GCAGATTAATGATGCAGTAATGCT 996
Db 203 Arglysphearglysglnserlyserileglnvalalmetglaasprvalthrhival 222
Qy 997 GCTATAGACTACAGCTGCTGAATTTATGAACTGTGAAGTATTTTAAATGAAAGATATGAA 1056
Db 223 Alaargglutrrilleglnlytyrargvalalargserphegylgylalalatyrrglu 242
Qy 1057 GCACAGAGATTTGATTTGATTTTGAAGCATATGAGATGCTTCATTTGAAGATACCTCT 1116
Db 243 Glulysargpheleuaspralaserlsernthraplysglnleuauargmetthrlyls 262
Qy 1117 ACTCTGCTATGCTGAACCTTGTCTCAAGTCTTATTTTCACTGCTGTTTAAACAGCTATA 1176
Db 263 ThrghlalaVallytrthPrometleuglnleuValilleYrvalAlametalilleu 282
Qy 1177 ATGTGCTCGCCAGTCAGGGAATTTGTCAGAGTACCTTACTGTGGAGATCTGTATG 1236
Db 283 MetpheleuValleutrp-----leuargglyaspralaserlglasprleuValala 300
Qy 1237 GTGATGAGATGCTTTTTCAGCTTTTACCTTACCTTGAACCTTGTGGAGATCTGTATG 1296
Db 301 TyrtlethrralaaaglyleuLeuProlysprollaarglnleuSerghlualaser 320
Qy 1297 GAGACTAGACAGACCTCATGATATGAACACTTGTCTTACTCTACCTCAAGTACAGACC 1356
Db 321 ThrvalGlnargglyValalaglyalaglnserllepheglnleuLeuAsprlualala 340
Qy 1357 CAA-----ATTAAACAAAGATGATGCATCTCCCTTCAGATACACCA 1401
Db 341 GlulguarpglnlythrhValglnlysglnargvalaserglyargleuglnvalarg 359
Qy 1402 CAGACAGTACCGGCTTGATGATGATGATTTGAAATCATTTGAGAGGCGCAGAA--- 1458
Db 360 -----AsnleuserphearglytrproglythrAsprylgln 371
Qy 1459 GTCCTTAGTGAATATCCTTTGAAATCCCTCAGAGAAAGAGTGCATGTAGAGAGT 1518
Db 372 ValleuaspralieserpheillealaglnproglylmetillealeuValalarg 391
Qy 1519 AGTGGTCAAGGAAAGACATATGAGCTTATTTTCCCTTCTTACGCTCAAAAG 1578
Db 392 SerllyserglylserthrleuAlasnleuValalproargpethyrglnhlaaspr 411
Qy 1579 GGTAGCAATTTATCTTGCTGCTGCAAAATATACAGATGAGCTGCAAGAGCTTCGAGG 1638
Db 412 GlulyslleleuLeuasprglyValalaglnasprlyrargleuargAsnleuarg 431
Qy 1639 GCAGTGGAGTGTACCTCAGAGATGCTGCTCTTCCATAATTAATTTATTAACATCTC 1698
Db 432 HistlealeuValThnglnlnvalThreuhheasnsprserValalAsnle 451
Qy 1699 TTATATGGAACATCATGCTTCACCT--GAGAAAGTATGCAAGTGGCAAAATTAAGT 1755
Db 452 AlatyrglyAsprleuAlaglyAlalproarggluglnilleglnargAlalalala 471
Qy 1756 GGACTTCATGATGCAATCTTCGAATGCCACATGATATGACCCCAAGTGGGGAAGA 1815
Db 472 AsnAlalysgluPheilleasprasnleuProglnglyPheasprthrglnuValgln 491
Qy 1816 GGAATCAAGCTTTTCAAGAGAGAAAGCAAGAGATGCAATTTGCAAGACCATTTTGAAG 1875
Db 492 GllyalaleuSerllyglyglnarglnrgleuAlalealalaglnleuLeuyls 511
Qy 1876 GACCCCAAGTCACTATGATGAAGCTTCTCATCGTTAGATTGATTAATGAAAGAG 1935
Db 512 AsprAlaproleuLeuilleuasprglualalThrserralaleuasprthrglnuserr 531
Qy 1936 ACTTTTCTTGTCGCAAGAGATGTGTCAAACAGAAATTTCTATTTTTCATTTGCAAC 1995
Db 532 HistleglnAlalaleuasprglualalmetylsglyargthrrhLeuValillealhis 551

```

```

Qy 1996 AGATTGCAACAGTGTGATGCAATGAAATCATTTGATCAGGTAAGTACCC 2055
Db 552 ArgleuserthrilleglnlysalalAsprleuilleuValmetasprglnlnleval 571
Qy 2056 GAACGTGTACCCACCATGTTTGTCTGCT--AACCTCAT 2094
Db 572 GluargglyserhlsalaglnleuLeualaglnasnlyhls 585

RESULT 13
US-09-749-340-6
; Sequence 6, Application US/09749340
; Patent No. US20020037920A1
; GENERAL INFORMATION:
; APPLICANT: Prichard, Roger K.
; APPLICANT: Xu, Ming
; APPLICANT: Ribeiro, Ana Paula
; APPLICANT: Blackhall, William J.
; APPLICANT: Beech, Rodin N.
; APPLICANT: Molento, Marcelo
; APPLICANT: Liu, Hao Yuan
; TITLE OF INVENTION: Methods for Detecting and Reversing
; TITLE OF INVENTION: Resistance to Macrocyclic Lactone Compounds
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products Corporation
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: United States
; ZIP: 07054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/749,340
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/067,676
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Renda, Barbara L.
; REGISTRATION NUMBER: 27,626
; REFERENCE/DOCKET NUMBER: 33,333-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 973-683-2153
; TELEFAX: 973-683-4109
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-749-340-6

Alignment Scores:
Pred. No.: 5,65e-47 Length: 1275
Score: 615.50 Matches: 196
Percent Similarity: 48.01% Conservative: 130
Best local Similarity: 26.87% Mismatches: 267
Query Match: 14,47% Indels: 86
DB: 10 Gaps: 21

AF133659 (1-2345) x US-09-749-340-6 (1-1275)
Qy 228 GTTCTTAGATGCTGCAAGAGCTCTCCAGATAGCCATGATGAAAGAGCATGTTG 287
Db 631 ValleuargglnlysglnlyglnuValleuasprserAspralaglnserAsprval 650
Qy 288 GCATGTATGACAGAGAGACTCCACACAGACCCAAAGAGGGTTAAAGATGTTGA 347

```

```

Db      651 -----SerProasp1lealaleuProh1leuSerSerleuAArgSerArgys- Gluse 668
Oy      348 TACTGGGAAAAATCATAAAGCAATGCTTTCTAT---GCTGGCCCAAGAACAGCCGCA 404
Db      668 rThrrArgSerAla1leSerAlaValProSerValArgSerMetGln1leGluMetGluas 688
Oy      405 TCTAGAGCTAGAGTTGCC----- 423
Db      688 pleuArgAlaLysProThrProMetSerLys1lePheTyrPheAsnArgAspLysTrpGl 708
Oy      424 ----ATTGCGTGGATTTTGGGTGGTGCAAGGCCATGAATATTGGTTCCTTCAT 479
Db      708 YTyrrPhe1leLeuGlyLeu1leAlaCys1le1leThrGlyThrVal1ThrProThrPheAl 728
Oy      480 GTTAAATATGCT-----GTAGACAGCCCTCAACCCAGATGTCGGGAA 521
Db      728 aValLeuTyrrAlaGln1le1leGlnValTyrrSerGluProValAspGlnMetLysGlyAl 748
Oy      522 CATGCTGAACCTGATGATGACACCAAAATACAGTTGCAACCATGGCAACAGCATTCGAT 581
Db      748 eValLeuPheTyrPcySgLyAla-----Phe1l 757
Oy      582 TGGCTATGCTGATACAGAGCTGAGCTGCTTTTAAAGATTGCAATGCAATG 641
Db      757 eVal1leGlyLeuVal1leAla1lePheAlaPhePheSerAla1le-----Cys1e 774
Oy      642 TGGCAAGGAGCCCAAGATTCATCCGAAGATAGCCAAAATGCTTTCTTCATCTTCA 701
Db      774 uGlyArgCysGlyAla1leLeuThrLysLeuArgPheGluAlaPheLysAsnLeu 794
Oy      702 CAACCTGATGCTGGTTTTCAC-----CTGACGACAGACGAGGAGCTTTATCTTAAGGC 755
Db      794 uArgGlnAsnVal1leGlyPheTyrAspAsp1leArgH1sg1lyThrGlyLysLeuCythrAr 814
Oy      756 TATTGACAGAGGAACAAGGGGTATCATAGTTTGTCTGAGTGGTTGGTATTTAATCTTCT 815
Db      814 gPheAlaThrAspAlaProAsnValArgTyr-----ValPheThrArg1e 829
Oy      816 TCCC---ATCATGTTTGAAGTATGCTGTCTGCTGCT-----GTTTGTATACAAAG 866
Db      829 uProGlyVal1leuSerSerValValThr1le1leGlyAlaLeuVal1leGlyPhe1lePh 849
Oy      867 CGGTGCCAGCTTGTCTTGGTATACCTTGA-----ACACTGTGATACATCAC 914
Db      849 eGlyTrrGlnLeuAlaLeu1leLeuMetValMetValProLeu1le1leGlySer----- 867
Oy      915 AGCATTCACAGTTGACATGACACGCTGAGAACTAGATTAGAAATGAATGAACAAGC 974
Db      868 -----GlyTyrPheGlnuMetArgMetGlnPheGlyLysLysMetLArgAspTh 883
Oy      975 AGAT-----AATGATGCGATATATGCTCTATAGACTACAGCTCGAATATGAAGACT 1028
Db      883 rGluLeuLeuGlnGluAlaGlyLysValAlaSerGlnAlaVal1leGluAsn1leArgThrVa 903
Oy      1029 GAAGTATTTTAAATGAAGATATGAAGACACAGATATGATGATTTTGAAGACGTA 1088
Db      903 lHisAlaLeuAsnArgGlnGluGlnPheHisPheMetTyrCysGluTyrrLeuLysGluPr 923
Oy      1089 TGAGACTGCTTATGAAAAAGTACTCTACTG---GCTATGCTGAACCTTTGGTCAAG 1145
Db      923 cTyrrArgLysAsnLeuCysGlnAlaHisThrTyrrGlyValPheAlaPheSerGlnSe 943
Oy      1146 TGCTATTTTCACTGCTGTTTAAAGCTATATAGTGTCTGCCAGTCAAGGAAATGTGC 1205
Db      943 rLeuLeuPhePheMetTyrAlaValAlaPheTrp1le----- 955
Oy      1206 AGGTACCTTACTGTTGGAGATCTAGTATAGTGAATGAGTGTCTTTACGCTTTCAT 1265
Db      956 -GlyAla1lePheValAspAsnHisSerMetGlnPro1leAspValTyrrArgValPhePh 975
Oy      1266 ACCCTGAACCTTTCTGGAACTGTATATAGAGACTAGACAAGCAAGCATATGATG-- 1323

```

```

Db      975 eAlaPheMetPheCysGlyGlnMetVal1leGlyAsn1leSerSerPhe1leProAspValVa 995
Oy      1324 -----AACACTTGTATTACTTACTACTACAGGTAGACACCAATTA 1364
Db      995 lLysAlaArgLeuAla1leAserLeuLeuPheTyrLeu1leGlnHisProSerGlu1leAs 1015
Oy      1365 AGACAAAGTATGATGCAATCTCCCTTCAGATCACACCAAG---ACAGCTACCGTGCCTT 1421
Db      1015 pAsn-----LeuSerGluAspGlyVal1ThrLysLys1leSerGlyHis1leSerPh 1032
Oy      1422 TGATTAATGCTATTTGATATCATTTAGAGGCCAG-----AAAGCTCTTACTGGAATATC 1475
Db      1032 eArgAsnValTyrrPheAsnTyrrProThrArgArgGln1leArgValLeuArgGlyLeuAs 1052
Oy      1476 CTTGAAGTCCCTCAGAAAGAAAGTGGCCATGTAGAGAGTGTGGTCAAGGAAAG 1535
Db      1052 nLeuGln1leAsnProGlyThrThrValAlaLeuVal1leGlnSerGlyCysGlyLysSe 1072
Oy      1536 CACAATAGTGAAGCTATTATTGCTTCTATAGCTTCMAAAGGGTAGCATTTATCTTGC 1595
Db      1072 rThrValMetAlaLeuLeuGlnuArgPheTyrrAsnGlnAsnLysGlyVal1leThrValAs 1092
Oy      1596 TGTCTCAAAATATATACAGATGAGCCTCGAAAGCCTCGAGGCGCAGTGGAGTGTACC 1655
Db      1092 pGlyGlnAsn1leArgAsnMetAsn1leArgAsnLeuArgGlnuVal1leCys1leVal1Se 1112
Oy      1656 TCAGATGCTGCTCTCTCCATTAATACTATTATTATTAACCTCTTATATGGAACATCAG 1715
Db      1112 rGlnGluProThrLeuPheAspCysThr1leMetGlnAsn1leCysTyrrGlyLeuAspAs 1132
Oy      1716 TGCTTCACT-----GAGAAAGTGTATGACGTGGCAAAATATAGCTGACTCATGTGC 1769
Db      1132 pProLysProSerTyrrGlnGlnVal1leAla1leAla1leAla1leMetAlaAsn1leHisAsnPh 1152
Oy      1770 AATCTTGAATGCGCACATGATATGACACCAAGTACGGAAGCAAGCACTCAAGCTTTC 1829
Db      1152 eValLeuGlyLeuLeuProGlnGlyTyrrAspThrArgVal1leGlyLysGlyThrGlnLeuSe 1172
Oy      1830 AGGAGAGAAAGCAAGAGAGTGAATTCGAAGAGCCATTTTGAAGACCCCGCAT 1889
Db      1172 rGlyGlyGlnLysGlnuArg1leAla1leAla1leArgAlaLeu1leLeuArgAspProProl1le 1192
Oy      1890 ACTGTATGATGAAGCTACTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1949
Db      1192 uLeuLeuAspGlnuAlaThrSerAlaLeuAspThrGluSerGlnuLys1leVal1leGlnAspAl 1212
Oy      1950 CATGAAGATGTGTCAAACACAGAACTTCTATTTTATGTCACACAGATGTCAACAGT 2009
Db      1212 aLeuGlnuValAlaArgGlnGlyArgThrCysLeuVal1leAlaHisArgLeuSerThr1l 1232
Oy      2010 GGTGATGACAGATGAATCATTTGCTTGGATCAGGTAAGGTAAAGTACCGAAGCTGTATCCCA 2069
Db      1232 eGlnAspSerAspVal1leVal1leMet1leGlnGlnGlyValAlaThrAspArgGlyThrHi 1252
Oy      2070 CCATGTTGCTTGTCTTAACCTCATATCTATTCATTCAGAAATGTGGACATACAG 2124
Db      1252 eGluHisLeuLeuMetLys---AsnAspLeuTyrrLysArgLeuCysGlnuThrGln 1269

```

```

RESULT 14
US-09-738-626-4751
; Sequence 4751, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHITO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO

```



```

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIORITY FILING DATE: 2000-12-18
; PRIORITY FILING DATE: 2000-12-18
; PRIORITY FILING DATE: 1999-12-16
; PRIORITY FILING DATE: 1999-12-16
; PRIORITY FILING DATE: 2000-04-07
; PRIORITY FILING DATE: 2000-04-07
; PRIORITY FILING DATE: 2000-08-03
; PRIORITY FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4751
; LENGTH: 1247
; TYPE: PR1
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-4751

Alignment Scores:
Pred. No.: 5,33e-45 Length: 1247
Score: 593.50 Matches: 167
Percent Similarity: 48.43% Conservative: 111
Best Local Similarity: 29.09% Mismatches: 265
Query Match: 13.95% Indels: 31
Gaps: 11

AF133659 (1-2345) x US-09-738-626-4751 (1-1247)
QY 418 GTTCCATTTGGCTGGGATTTTGGTGGTGCAGGATGATATTTGGTCCCTTC 477
DB 695 ValAlaValAlIleAlaLeuLeuValGlyVal--AlaAlaLeuLeuAlaPheProthr 713
QY 478 ATGTTTAATATGCTGTAGACAGCTCAACAGATGTCGGAACATGCTGACCTGAT 537
DB 714 LeuMetArgAlaAlaIleLeuPheSerGlyValGln-----Ala 725
QY 538 GATGACCAAAATATACATGTTGACCAACGAGACAGATGTCGATGGCTGATGATCA 597
DB 726 GlnSerThrSerThrLeuTrpTrpIleAlaIleAlaGlySerValValLeuLeuSer 745
QY 598 AGACCTGGAGCTGCTTTTAAAGAACTTGAATGACATATTTGGCAAGTACCCAG 657
DB 746 TrpAlaAlaAlaAla-----IleAsnThrIleIleThrAlaArgThrGlyGlu 761
QY 658 AATCAATCCGAGAGATAGCAAAATGCTTTCCTCATCTTCAACATGCTGATCGGT 717
DB 762 ArgLeuLeuTrpGlyLeuArgLeuArgSerPheValIhSleuLeuArgLeuSerMetSer 781
QY 718 TTTCACCTGACAGACAGACGGAGCTTTATCTAAGCTATTTGACAGAGAAAGGGGT 777
DB 782 TyrPheGluArgThrMetSerGly-----ArgIleMetThrArgMetThrThrAsp 798
QY 778 ATCAGTTTGCTGAGTCTTTGGTATTTATCTTCTCCATCATGTTGAAAGTATG 837
DB 799 IleAspAsnLeuSerSerPheLeuGlnSerGlyLeuAlaGlnThrValValSerValGly 818
QY 838 CTTCGAGTGGTGTGTTGTTGATTAACAATGC-----GGTGGCCAGTTTGGTGAAC 891
DB 819 ThrIleuIleGlyValValThrMetLeuAlaIleThrAspAlaGlnLeuAlaLeuValAla 838
QY 892 CTTCGAAACACTGTGTATACACAGACATTCACAGTTCGACGACGAGTGGAGAACTAGA 951
DB 839 LeuSerValValProIleIleIleValLeuThrLeuIlePheArgArgIleSerSerArg 858
QY 952 TTTGAATAGAAATGACAAAGCAGATATATGATGACGATATGCTGCTATGACTCATG 1011
DB 859 LeuTrpThrAlaSerArgGlnGlnAlaSerGlnValAsnAlaValPheHisGlnSerIle 878
QY 1012 CTGAATTATGAACTGTGAGATATTTTAATAGAAAGATATGAGACAGACAGATATGAT 1071
DB 879 AlaGlyLeuAlaGlnThrAlaGlnMetHisArgMetGlnSerGlnValPheAspAsnTrpAla 898
QY 1072 GGATTTTGAAGAGTATGAGACTGCTTCATTTGAAAAGTACCTCTACCTGCTATGCTG 1131
```

```

DB 899 GlyGluAlaGlnGluPheArgArgLeuArgValIysSerGlnThrAlaIleAlaIleTrp 918
QY 1132 AACTTTGTCMAAGTGTATTTTTCAGTGTGCTTTTAAACAGTAAATGTCGTCAGC 1191
DB 919 PheProGlyLeuGlyAlaLeuSerGlnIleAlaGlnAlaLeuValLeuGlyPheGlyAla 938
QY 1192 CAGGAAATGTGTGACAGTACCTTACTGTTGAGATCTA-----GTAAATGTGAAT 1242
DB 939 LeuGlnValThrArgGlyAspIleSerThrGlyValLeuValAlaPheValLeuTyrMet 958
QY 1243 GAGTGTCTTTTTCAGCTTTTCAATTAACCTGACCTTTCGGGAAGTGTATATGAGACT 1302
DB 959 GlyLeuMetPheGly-----ProIleGlnGlnLeuSerGlnIlePheAspSerTrp 975
QY 1303 AGACAAACCTCATATGATATGAACACCTGTGTTACTCTACCTCAAGATGACCAACCAAT 1362
DB 976 GlnGlnAlaAlaValGlyPheArgArgIleThrGlnLeuLeuAlaThrGlnProSerVal 995
QY 1363 AAAGACAAAGTATGAGATGTCATCCCTTCAGATCACACAGACAGACAGTACC-----GTG 1416
DB 996 ProAsp-----LeuGlyThrAsnArgHisAlaArgGlnAlaIleThrGlnProLeu 1012
QY 1417 GCCTTGATATATGATGCTTTTGAATATGATGAGGCGCAAGATCCTTATGCAATATCC 1476
DB 1013 LeuLeuAspAspValThrPheGlyTyrSerAsp---AspProIleLeuAspAsnValThr 1031
QY 1477 TTTGAATCCCTGACGAGAAAGAAAGATGTCATGTCAGAGTATGAGTGTGAGAAAGC 1536
DB 1032 ValGlnIleGluProGlyThrThrValAlaValGlyProThrGlyAlaGlyLysSer 1051
QY 1537 ACAATATGAGAGCTATATTTCCCTTCATATGATGCTCAAAAGGATGACATTTATCTTGT 1596
DB 1052 ThrValValLysLeuLeuSerArgLeuTrpAspProAsnGlnGlyAlaValLysAlaGly 1071
QY 1597 GGTCAAAATATATACAGATGTCAGCTGAGGAAAGCCTTGAGGCGCAGTGGAGTGTACT 1656
DB 1072 GlnIleAspIleLysAspPheProThrAlaAspTrpArgArgThrIleGlyThrValPro 1091
QY 1657 CAGAGTCTGCTCCCTTCATATATGATTTTAAACACCTTATTTGAAACATCAGT 1716
DB 1092 GlnGlnAlaIhSleuPheSerGlySerIleAlaAspAsnIleGlyTrpGlyCysArgGlu 1111
QY 1717 GCTTCACTGAGAGAGTATGAGTGTGCAAAATGATGCTGACTTCTGATGATGCAATCTT 1776
DB 1112 AlaSerThrSerLysIleGlnAlaAlaAlaArgArgValGlyAlaLeuAlaAlaIleAla 1131
QY 1777 GCAATGCCATGATATGACACCCAAAGTAGGGAACGAGAGCTCAAGCTTTTCAGAGGA 1836
DB 1132 AlaIleProAspGlyPheAsnHisGlnValGlyGlnArgGlyArgAsnLeuSerSerGly 1151
QY 1837 GAAAGCAAGAGATGAGATATGCAAGGCCATTTTGAAGAACCCCGACATATCTAT 1896
DB 1152 GlnArgGlnLeuIleAlaLeuAlaArgAlaGlnLeuIleGlnProSerIleMetLeuLeu 1171
QY 1897 GATGAGCTACTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1956
DB 1172 AspGlnAlaThrSerThrLeuAspProAlaThrGlnAlaValIleLeuAsnAlaSerAsp 1191
QY 1957 GATGTGTCAACACAGACAACTTATTTTTCATTTGACACAGATTTGTCAGAGTGTGAT 2016
DB 1192 ArgValThrLysGlyArgThrSerIleIleValAlaHisArgLeuAlaThrAlaLysArg 1211
QY 2017 GCAGATGAATATCATGTGCTTGTGATCAGGGTAAAGTATGAGTATGATGATGATGATG 2076
DB 1212 AlaAspArgIleLeuValValGlnGlnGlyArgIleIleGlnAspGlySerHisAspAla 1231
QY 2077 TTGCTGTCAACCTCATATGATATGATGATGATGATGATGATGATGATGATGATGATG 2118
DB 1232 LeuLeu---SerAlaAsnGlyThrTrpAlaArgMetTrpHis 1244

RESULT 15
US-09-873-409-6
```



Db 574 ThrTyrSerThrGluArgLysThrAsnSerLeuProLeuHisSerValLysSerIleLys 593

Search completed: February 15, 2003, 06:10:50  
Job time : 40.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2003, 05:54:00 ; Search time 20.5 Seconds  
(without alignments)  
6731.390 Million cell updates/sec

Title: AF133659  
Perfect score: 4254  
Sequence: 1 ATGGCCGCTGCTCGCATGCA.....TATTAAAAAATCATACATT 2345

Scoring table:  
BLOSUM62 Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODE=frame+n2p,model -DEV=x1h  
-O=/cgn2\_1/USPTO\_spool/AF133655/runat\_10022003\_155441\_29551/app\_query.fasta\_1.2503  
-DB=Issued\_Patents\_AA -OPMT=faetan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsun62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTEXT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRA=AF133659 @CGN 1.1 18 @runat\_10022003\_155441\_29551 -NCPU=6 -ICPU=3  
-NO\_XLXPY -NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-MAFN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCPUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description         |
|------------|--------|-------------|--------|-------|---------------------|
| 1          | 3809   | 89.5        | 747    | 2     | US-08-895-522-1     |
| 2          | 3809   | 89.5        | 747    | 3     | US-09-195-391-1     |
| 3          | 3342   | 78.6        | 694    | 2     | US-08-895-522-3     |
| 4          | 3342   | 78.6        | 694    | 3     | US-08-895-522-3     |
| 5          | 1516.5 | 35.6        | 694    | 2     | US-08-895-522-4     |
| 6          | 1516.5 | 35.6        | 694    | 2     | US-08-895-522-4     |
| 7          | 1201.5 | 28.2        | 475    | 3     | US-09-195-391-4     |
| 8          | 695.5  | 16.3        | 582    | 4     | US-09-134-001C-4730 |
| 9          | 641.5  | 15.1        | 707    | 4     | US-08-772-270A-4    |
| 10         | 637.5  | 15.0        | 580    | 4     | US-09-134-001C-5611 |
| 11         | 618.5  | 14.5        | 711    | 3     | US-08-772-270A-12   |
| 12         | 594.5  | 14.0        | 791    | 1     | US-08-394-880B-2    |

|    |       |      |      |   |                   |                    |
|----|-------|------|------|---|-------------------|--------------------|
| 13 | 583.5 | 13.7 | 1275 | 3 | US-09-120-513-2   | Sequence 2, App1   |
| 14 | 583.5 | 13.7 | 1275 | 4 | US-09-450-105-2   | Sequence 2, App1   |
| 15 | 583   | 13.7 | 1279 | 2 | US-08-784-649A-2  | Sequence 2, App1   |
| 16 | 578.5 | 13.6 | 1280 | 6 | US-08-583-276-19  | Sequence 19, App1  |
| 17 | 578.5 | 13.6 | 1280 | 6 | 5206352-4         | Patent No. 5206352 |
| 18 | 565.5 | 13.3 | 1280 | 2 | US-08-752-447-2   | Sequence 2, App1   |
| 19 | 565.5 | 13.3 | 1280 | 4 | US-09-316-167-2   | Sequence 2, App1   |
| 20 | 562.5 | 13.2 | 1334 | 2 | US-08-996-545-2   | Sequence 2, App1   |
| 21 | 562.5 | 13.2 | 1334 | 4 | US-09-328-320-2   | Sequence 2, App1   |
| 22 | 562.5 | 13.2 | 717  | 4 | US-08-924-629C-5  | Sequence 5, App1   |
| 23 | 554   | 13.0 | 1408 | 1 | US-08-612-521-2   | Sequence 2, App1   |
| 24 | 546   | 12.8 | 1307 | 1 | US-08-395-246C-2  | Sequence 2, App1   |
| 25 | 545.5 | 12.8 | 1349 | 2 | US-08-612-734B-2  | Sequence 2, App1   |
| 26 | 536   | 12.6 | 1308 | 2 | US-08-996-644-2   | Sequence 2, App1   |
| 27 | 536   | 12.6 | 1308 | 3 | US-09-352-552-2   | Sequence 2, App1   |
| 28 | 518.5 | 12.2 | 748  | 4 | US-09-061-764A-19 | Sequence 19, App1  |
| 29 | 513   | 12.1 | 400  | 4 | US-08-961-083-190 | Sequence 190, App  |
| 30 | 509   | 12.0 | 1263 | 4 | US-09-351-224E-11 | Sequence 11, App1  |
| 31 | 507.5 | 11.9 | 1302 | 1 | US-08-332-537-2   | Sequence 2, App1   |
| 32 | 470.5 | 11.1 | 686  | 4 | US-09-061-764A-15 | Sequence 15, App1  |
| 33 | 442   | 10.4 | 1437 | 3 | US-09-061-400-2   | Sequence 2, App1   |
| 34 | 442   | 10.4 | 1453 | 2 | US-09-001-273-2   | Sequence 2, App1   |
| 35 | 442   | 10.4 | 1453 | 4 | US-08-843-459A-2  | Sequence 2, App1   |
| 36 | 428   | 10.1 | 1622 | 4 | US-08-972-927-6   | Sequence 6, App1   |
| 37 | 421   | 9.9  | 1621 | 4 | US-08-972-927-3   | Sequence 6, App1   |
| 38 | 414.5 | 9.7  | 653  | 4 | US-09-061-764A-2  | Sequence 2, App1   |
| 39 | 408   | 9.6  | 1261 | 4 | US-09-605-785-538 | Sequence 538, App  |
| 40 | 408   | 9.6  | 1261 | 4 | US-09-439-313-538 | Sequence 538, App  |
| 41 | 406   | 9.5  | 327  | 1 | US-08-463-092B-9  | Sequence 9, App1   |
| 42 | 406   | 9.5  | 327  | 2 | US-08-460-907B-9  | Sequence 9, App1   |
| 43 | 401.5 | 9.4  | 1531 | 1 | US-08-141-893-2   | Sequence 2, App1   |
| 44 | 401.5 | 9.4  | 1531 | 1 | US-08-463-092B-2  | Sequence 2, App1   |
| 45 | 401.5 | 9.4  | 1531 | 1 | US-08-463-092B-4  | Sequence 4, App1   |

## ALIGNMENTS

RESULT 1  
US-08-895-522-1  
; Sequence 1, Application US/08895522  
; Patent No. 5858719  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Shah, Puryi  
; TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE  
; TITLE OF INVENTION: TRANSPORT PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSER: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/895,522  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: Pf-0336 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555

```
/ TELEFAX: 415-845-4166
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 747 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: OVARNOT02
/ CLONE: 545981
/ US-08-895-522-1

Alignment Scores:
Pred. No.: 0 Length: 747
Score: 3809.00 Matches: 746
Percent Similarity: 99.87% Conservative: 0
Best Local Similarity: 99.87% Mismatches: 1
Query Match: 89.54% Indels: 0
DB: Gaps: 0

AF133659 (1-2345) x US-08-895-522-1 (1-747)

QY 16 ATGCATTCCTGGCGCTGGCGCGCGCGCGCTCTTGGAAAAGCCGCGACCTCCGCG 75
Db 1 MetHisSerTPrArgTrpAlaIaIaIaIaIaIaIaIaPheGluLysArgTrgHisSerAla 20

QY 76 ATTCGATCCCGCCCTTTAGTCTCTGTTAGCGGCTCAGGTCGCCAGTGGAGGCGACATCA 135
Db 21 IleuIleuIleArgProLeuValSerValSerGlySerGlyProGlnTrpArgProHisGln 40

QY 136 CTGCGCGCTTGGGAACCGCTCAGACCTTACAGATTCAGAGTCTGTAATAAAGTATTCACA 195
Db 41 LeuGluAlaLeuGluThrlAlaArgAlaTyrglnIleProGlnSerIleuLysSerIleThr 60

QY 196 TGGCAGAGATTGGGAAAAGGCAATTCAAGACAGTTCTTAGATGCTGCAAAAGGCTCTCCAG 255
Db 61 TrpGlnArgLeuGluLysGluYasnSerGlyGlnPheLeuAspAlaAlaLysAlaLeuGln 80

QY 256 GATATGGCCCATGATAGAAAAGAGACATGTTGGCATGTGTCATGCGAGGAGAGACTCCAC 315
Db 81 ValTrpProLeuIleGluLysArgTrgThrCysTrpHisGlyHisAlaGluLysLeuHis 100

QY 316 ACAGACCCCAAGAAAGAGGTTAAAGATGTTGATACCTCGGAAATGATAAAGCAATGCTT 375
Db 101 ThrAspProLysGluGluLysAspValAspThrArgLysIleIleLysAlaLysLeu 120

QY 376 TCTTATGTGTGGCCCAAGACAGGCGCAGATCTACAGAGCTAGAGTTGCCATTTGCGTGGGA 435
Db 121 SerTyrrValTrpProLysAspArgProAspLeuArgAlaArgValAlaIleSerLeuGly 140

QY 436 TTTTGGGTGGTGCAAAAGGCCCATGAATATTTGGTTCCTTCAATGTTTAAATATGCTGTGA 495
Db 141 PheLeuGluGlyAlaLysAlaMetAsnIleValIleProPheMetPheLysTrpIleVal 160

QY 496 GACAGCCTCAACAGATGTCGGGAAACATGTGTAACCTGAGATGACCAATATCAGTT 555
Db 161 AspSerLeuAsnGlnMetSerGlyAsnMetLeuAsnLeuSerAspAlaProAsnThrVal 180

QY 556 GCAACCATGCAACAGACAGTTCGATGTGGTATGCTGATCAAGAGCTGAGACTGCTTTT 615
Db 181 AlaThrMetAlaThrAlaValIleuIleGlyTyrglyValSerArgAlaGluAlaIaPhe 200

QY 616 TTTTACGAGAGTTGCAAAATCAGATTTTGGCAAGTACGCCAAGATTCATCCGAAGAATA 675
Db 201 PheAsnGluValArgAsnAlaValIlePheGlyLysValAlaGlnAsnSerIleArgTrgIle 220

QY 676 GCCAAATAATGTTTCTCATCTTCAACAGCTGAGATCTGGGTTTCACTGAGCAAGAG 735
Db 221 AlaLysAsnValIlePheLeuHisLeuHisAsnLeuAspLeuGlyPheHisLeuSerArgGln 240

QY 736 ACGGAGGCTTTATCTTAAGGCTATTGACAGAGGAACAAGGAGGTATGAGTTTGTCTCGAGT 795
Db 736 ACGGAGGCTTTATCTTAAGGCTATTGACAGAGGAACAAGGAGGTATGAGTTTGTCTCGAGT 795
```

```
Db 241 ThrGluAlaLeuSerLysAlaIleAspArgGlyThrArgGlyLysSerPheValLeuSer 260
QY 796 GCTTGTGATTTAATCTTCTCCATCATGTTTGAAGGATGCTTGCAGTGAGTGTG 855
Db 261 AlaLeuValPheAsnLeuLeuProIleMetPheGluValMetLeuValSerGlyValLeu 280
QY 856 TATTACAAATCCGGTGCCCATGTTGCTTGGTAAACCTTGGACACTTGATACATACACA 915
Db 281 TyrrTyrrLysCysGlyAlaGlnPheAlaLeuValThrLeuGlyThrLeuGlyThrTyrrThr 300
QY 916 GCATTCAACAGTTGCACTCACCGGTGAGAACTGATTTAGAAATAGAAATGAACAAAGCA 975
Db 301 AlaPheThrValAlaValThrArgTrpArgThrArgPheArgIleGluMetAsnLysAla 320
QY 976 GATATATGATGAGGAATATGCTGCTATGACTCACCTGCGTAATTAAGACGTGAAGTAT 1035
Db 321 AspAsnAspAlaGlyAsnAlaAlaIleAspSerLeuLeuAsnTyrrGluThrValLysTyrr 340
QY 1036 TTTATATATGAAAAGATATGAAGCACAGAGATATGATGATTTTGAAGACGTATGAGACT 1095
Db 341 PheAsnAsnGluArgTyrglyAlaIleGlnArgTyrrAspGlyPheLeuLysThrTyrrGluThr 360
QY 1096 GCTTCATTGAAAAGTACTCTTACTCTGTGCTATGCTGAACCTTGGTCAAGTCTATTATTC 1155
Db 361 AlaSerLeuLysSerThrSerThrLeuAlaMetLeuAsnPheGlyGlnSerAlaIlePhe 380
QY 1156 AGTGTGCGTTTACAGCATATATATGCTGCGCAGTCAAGGAAATTGCGCAGTACCTT 1215
Db 381 SerValGlyLeuThrAlaIleMetValLeuAlaSerGlnGlyIleValAlaGlyThrLeu 400
QY 1216 ACTGTGGAGATCTAGTAATGATGTAATGACTGCTTTTTCAGCTTTCATTAACCCCTGAAC 1275
Db 401 ThrValGlyAspLeuValMetValAsnGlyLeuLeuPheGlnLeuSerLeuProLeuAsn 420
QY 1276 TTTCTGGGAACTGATATATAGAGACATAGACACACTCATATATATGAAACCTTGT 1335
Db 421 PheLeuGlyThrValTyrrArgGluThrArgGlnAlaLeuIleAspMetAsnThrLeuPhe 440
QY 1336 ACTCTACTCAAGGTATAGACACCCCAATTAAGCAAGAGTATGAGCATCCCTTCAGATC 1395
Db 441 ThrLeuLeuLysValAspThrGlnIleLysAspLysValMetAlaSerProLeuGlnIle 460
QY 1396 ACACACACAGACAGCTACCGTGGCTTGTATATGATGCAATTTGAAATCATTTAGAGGCGCAG 1455
Db 461 ThrProGlnThrAlaThrValAlaPheAspAsnValHisPheGluTyrrIleGluGlyGln 480
QY 1456 AAAGCTCTTACTGGAATATCTCTTGAAGTCCCTCGAGGAAAGAAAGTGCCATTGTGGA 1515
Db 481 LysValLeuSerGlyIleSerPheGluValProIaGlyLysValAlaIleValAlaGly 500
QY 1516 GGTAGTGGTCAAGGAAAGCAAAATAGTGAGCTATTTATTTGGCTTCTATAGAGCTCAA 1575
Db 501 GlySerGlySerGlyLysSerThrIleValArgLeuLeuPheArgPheTyrrGluProGln 520
QY 1576 AAGGTATGACTTTTATCTTGTGCTCAAAATATATACAGATGAGCTCGAAGGCTTCGG 1635
Db 521 LysGlySerIleTyrrLeuAlaGlyGlnAsnIleGlnAspValSerLeuGluSerLeuArg 540
QY 1636 AGGCAAGTGGGAGTGTATCTCAAGATGCTGTCTCTTCCATTAATACTATTATTATTAAC 1695
Db 541 ArgAlaValGlyValValProGlnAspAlaValLeuPheHisAsnThrIleTyrrLysAsn 560
QY 1696 CTTTATATATGAAACATCATGAGCTTCACTCCTGAGGAAGGTATGAGTGGCAAAATTTAGCT 1755
Db 561 LeuLeuTyrrGlyAsnIleSerAlaSerProGluValTyrrAlaValAlaLysLeuAla 580
QY 1756 GGACTTCATGATGCAATCTTCCGAATGCGACATGATATGACACCCAAGTAGGGGAAAGCA 1815
Db 581 GlyLeuHisAspAlaIleLeuArgMetProHisGlyTyrrAspThrGlnValGlyGluArg 600
QY 1816 GGACTCAAGCTTTTCAAGAGGAAAGCAAAAGATGACCAATTTGCAAGAGCCATTTTGAAG 1875
Db 601 GlyLeuLysLeuSerGlyGlyGlyGluGlnArgValAlaIleAlaArgAlaIleLeuLys 620
```

QY 1876 GACCCCCAGTCACTCTGATGATGAAGCTACTTCATCGTTAGATTGCGATTCTGAAG 1935  
| | | | |  
Db 621 AspProValIleLeuYrAspGluAlaThrSerSerLeuAspSerIleThrGluGlu 640  
QY 1936 ACTTTCCTGGTGCATGAAGAGATGTGCTCAACACAGAACTTCTATTTCATTGCGAC 1995  
| | | | |  
Db 641 ThrIleuGluAlaMetLysAspValValLysHisArgThrSerIlePheIleAlaHis 660  
QY 1996 AGATTGTCACAGTGTGATGATGACATGAATTCATTGCTTGATCGAGGAGTAAAGTACC 2055  
| | | | |  
Db 661 ArgLeuSerThrValValAspAlaAspGluIleIleValIleuAspGlnGlyValAla 680  
QY 2056 GAACGTGATCCACCACTGTTGCTTGTACCCCTCATAGTATCTATTCAGAAATGTGG 2115  
| | | | |  
Db 681 GluArgGlyThrHisHisGlyLeuIleuAlaAsnProHisSerIleYrSerGluMetTyr 700  
QY 2116 CATTCACAGACGACCGCTGTGCAGACCATATTAACCCCAATGGAGCAAGAAAGAA 2175  
| | | | |  
Db 701 HisThrIleuSerArgValGlnAsnHisAspAsnProLysTyrGluAlaLysGlu 720  
QY 2176 AATATATCCAAAGAGAGAAAGAAAGAACTACAGAAATTTGCAATAGTGTGAA 2235  
| | | | |  
Db 721 AsnIleSerLysGluGluGluArgLysLeuGlnGluIleValAsnSerValLys 740  
QY 2236 GCGCTGTGAAACTGTGTGTGC 2256  
| | | | |  
Db 741 GlySerGlyAsnCySerCys 747  
RESULT 2  
US-09-195-391-1  
; Sequence 1, Application US/09195391  
; Patent No. 6080842  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Shah, Puri A.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE  
; TITLE OF INVENTION: TRANSPORT PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/195,391  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/895,522  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0336 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 747 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

IMMEDIATE SOURCE:  
; LIBRARY: OVARNOT02  
; CLONE: 545981  
US-09-195-391-1  
Alignment Scores:  
Pred. No.: 0  
Score: 3809.00  
Percent Similarity: 99.87%  
Best Local Similarity: 99.87%  
Query Match: 89.54%  
DB: 3  
Length: 747  
Matches: 746  
Conservative: 0  
Mismatch: 1  
Indels: 0  
Gaps: 0  
AF133659 (1-2345) x US-09-195-391-1 (1-747)  
QY 16 ATGCATCTTGGGCGCGGCGCGCGCGCGCTTTCGAAAAGCGCGCATCCGCG 75  
| | | | |  
Db 1 MetHisSerTyrArgTyrAlaAlaAlaAlaAlaAlaPheGluLysArgArgHisSerAla 20  
QY 76 ATTCTGATCCGCGCTTGTAGTCTCTGTAGCGGCTCAGGTCCGACGTGAGCGCACATCAA 135  
| | | | |  
Db 21 IleuIleuArgProLeuValSerValSerGlySerGlyProGlnTyrArgProHisGln 40  
QY 136 CTGCGCGCGCTTGGGAAACCGCTCAGCGCTACAGATTCCAGAGTCATTAAGATATACA 195  
| | | | |  
Db 41 LeuGlyAlaLeuGlyThrAlaArgAlaTyrGlnIleProGluSerLeuLysSerIleThr 60  
QY 196 TGGCAGAGATTGGGAAAAGGCAATTCAGGACAGTCTTATGCTGCAAGGCGCTTCAG 255  
| | | | |  
Db 61 TrpGlnArgLeuGlyLysGlyAsnSerGlyGlnPheLeuAspAlaAlaLysAlaLeuGln 80  
QY 256 GTATGCGCATGTAGTAAAGAGAGACATGTGGCATGTGTCATGTCAGAGAGAGACTTCAC 315  
| | | | |  
Db 81 ValTyrProLeuIleGluLysArgGlnCysTyrHisGlyHisAlaGlyGlyLeuHis 100  
QY 316 ACAGACCCAAAGAGAGGTTAAAGATGTTGATCTGCGAAATCATAAAGCAATGCTT 375  
| | | | |  
Db 101 ThrAspProLysGluGlyLeuLysAspValAspThrArgLysIleIleLysAlaLysLeu 120  
QY 376 TCTTATGTGTGCGCAAGAGCAGGCGCATCTCAGAGCTTACAGTTGCGATTTGGCTGG 435  
| | | | |  
Db 121 SerTyrValTyrProLysAspArgProAspLeuArgAlaArgAlaIleSerLeuGly 140  
QY 436 TTTTGGTGTTGTCGCAAGGCGCATGATATGTGTGCTTCCCTTCATGTTAAATAGCTGA 495  
| | | | |  
Db 141 PheLeuGlyGlyAlaLysAlaMetAsnIleValValProPheSerPheLysTyrAlaVal 160  
QY 496 GACAGCCTCAACGAGATGTGCGGAAAATGCTGAGCTGATGATGATGATGATGATGAT 555  
| | | | |  
Db 161 AspSerLeuAsnGlnMetSerGlyAsnMetLeuAsnLeuSerAspAlaProAsnThrVal 180  
QY 556 GCACACATGCGAACACAGATGTTGATGTTGCTATGTGTATCAAGAGCTGAGCTGCTTT 615  
| | | | |  
Db 181 AlaThrMetAlaThrAlaValLeuIleGlyTyrGlyValSerArgAlaGlyAlaAlaPhe 200  
QY 616 TTTACGAAATGTCGAAATGATGATTTGGCAGAGTACCGCAATTCATCCGAAAGATA 675  
| | | | |  
Db 201 PheAsnGluValArgAsnAlaValPheGlyLysValAlaGlnAsnSerIleArgGlnIle 220  
QY 676 GCCAAAATGTCTTCTCATCTTCACAACTTGATCTGGGTTTCACTGACGACAGACAG 735  
| | | | |  
Db 221 AlaLysAsnValPheLeuHisLeuHisAsnLeuAspLeuGlyPheHisLeuSerArgGln 240  
QY 736 ACGGAGCTTATCTTAAGCTATTTGACAGAGAAACAAGGGGTATCACTTTGCTCGACT 795  
| | | | |  
Db 241 ThrGlyAlaLeuSerLysAlaIleAspArgGlyThrArgGlyLysSerPheValLeuSer 260  
QY 796 GCTTTGATTTATCTTCTTCCATCATGTTGAAGTATGTTGATGATGATGATGATGATG 855  
| | | | |  
Db 261 AlaLeuValPheAsnLeuLeuProIleMetPheGlnValMetLeuValSerGlyValLeu 280  
QY 856 TATTACAAATGCGGTGCCAGATTGCTTTGTAACCTTGAACACTTGGTACATACACA 915  
| | | | |



Db 281 TyrTyrLysCyseGlyAlaGlnPheAlaLeuValThrLeuGlyThrLeuGlyThrTyrThr 300  
 QY 916 GGATTCACAGTGGACGTGCACAGGTGAGAGAACTAGATTGATGAATGAATGAACAAGCA 975  
 Db 301 AlaPheThrValAlaValAlaThrArgTrpArgTrpArgPheArgIleGluMetCysnLysAla 320  
 QY 976 GATTAATGATGACAGTAATGCTGCTATAGACTGACTGCTGGAATTAATGAACCTGTGAAGTAT 1035  
 Db 321 AspaAsnAspAlaGlyAsnAlaAlaIleAspSerLeuLeuAsnTyrGluThrValLysTyr 340  
 QY 1036 TTTATATATGAAGAATTTGACACACAGATATGATGATTTTGAAGAAGCTATGAGACT 1095  
 Db 341 PheAsnAsnGluArgTyrGluAlaGlnArgTyrAspGlyPheLeuLysThrTyrGluThr 360  
 QY 1096 GCTTATGTAAGAATGACCTCTACTGCTGCTATGTAACCTTTGGTCAAAAGTCTATTTTC 1155  
 Db 361 AlaSerLeuLysSerThrSerThrLeuAlaMetLeuAsnPheGlyGlnSerAlaIlePhe 380  
 QY 1156 AGTGTGGTTTAAACAGCTATMAATGGTGTGCTGCGCAGTCAAGGAATTTGTGACAGTACCCTT 1215  
 Db 381 SerValGlyLeuThrAlaIleMetValLeuAlaSerGlnGlyIleValAlaGlyThrLeu 400  
 QY 1216 ACTGTTGAGAGATCTTGTAATGCTGAATGAGACTGCTTTTCACTTTCACTTACCCTGTAAC 1275  
 Db 401 ThrValGlyAspLeuValMetValAsnGlyLeuLeuPheGlnLeuSerLeuProLeuAsn 420  
 QY 1276 TTTCTGGGGAACGTATATATGAGAGACTAGACAAAGCACTGATGATATGAACCTTGTTT 1335  
 Db 421 PheLeuGlyThrValTyrArgGluThrArgGlnAlaLeuIleAspPheCysnThrLeuPhe 440  
 QY 1336 ACTTACTCAAGGTAGACACACCCAAATTAAGACAAAGTATGGCACTCTCCCTTGACATC 1395  
 Db 441 ThrLeuLeuLysValAspThrGlnIleLysAspLysValMetAlaSerProLeuGlnIle 460  
 QY 1396 ACAACAACAGACGCTACCGTGGCTTTGATATGTCATTTTGAATPACATTGAGGCCAG 1455  
 Db 461 ThrProGlnThrAlaThrValAlaPheAspAsnValIleAspGluTyrIleGluGln 480  
 QY 1456 AAAAGCTTAGTGAATATCCTTTGAAGTCCCTGACAGGAAAGAAAGTGGCCATTGAGGA 1515  
 Db 481 LysValIleSerGlyIleSerPheGluValProIleGlyLysValAlaAlaIleValGly 500  
 QY 1516 GGTACTGGGTCAAGGAAAGCAAAATAGTAGAGCTATTATTTGCTTCTATGAGCCTCA 1575  
 Db 501 GlySerGlySerGlyLysSerThrIleValAlaGlyLeuLeuPheArgPheTyrGluProGln 520  
 QY 1576 AAGGTAGCATTTATCTGCTGCTGCAAAATATACAGATGTGAGCCTGGAAGCCTTGG 1635  
 Db 521 LysGlySerIleTyrIleuAlaGlyGlnAsnIleGlnAspValSerLeuGluSerLeuArg 540  
 QY 1636 AGGGAGTGGAGGTGATACCTGAGATGCTGCTCTTGCATATATCTATTATTAACAAC 1695  
 Db 541 ArgAlaValAlaGlyValAlaProGlnAspAlaValAlaLeuPheAsnThrIleTyrTrpAsn 560  
 QY 1696 CTCTTATATGGAACATCAGTCTTCACTGAGGAAGTATGACAGTGGCAAAATTAAGCT 1755  
 Db 561 LeuLeuTyrGlyAsnIleSerIleAspProGluGluValTyrAlaValAlaLysLeuAla 580  
 QY 1756 GGACTTCATGATGCAATTTTGAATGCCACATGGATATGACACCCAGTAGGGGAACGA 1815  
 Db 581 GlyLeuHisAspAlaIleLeuArgMetProHisGlyTyrAspThrGlnValGlyGluArg 600  
 QY 1816 GGAGCTCAAGCTTTCAGGAGGAGAAAGAAAGAGTAGCAATTCGAAGCCATTTTGAAG 1875  
 Db 601 GlyLeuLysLysSerGlyGlyLysGlnArgValAlaIleAlaArgAlaIleLeuLys 620  
 QY 1876 GACCCCAAGTCATCTCTATGATGAGAGTACTCATCTGATTCGATTCGATTAAGAG 1935  
 Db 621 AspProProValIleLeuLysArgGluAlaThrSerSerLeuAspSerIleThrGluGlu 640  
 QY 1936 ACTATTTTGGTGCATGAAGAGTGTGCAAAACACAGAACTTCTATTTCATTGCAAC 1995  
 Db 641 ThrIleLeuGlyAlaMetLysAspValValLysHisArgThrSerIlePheIleAlaHis 660

QY 1996 AGATTGTCACAGTGGTTGATGACATGAATCATTTGTTGATCAGGGTAAAGTACCC 2055  
 Db 661 ArgLeuSerThrValValAspAlaAspGluIleIleValLeuAspGlnGlyLysValAla 680  
 QY 2056 GAACGTGATCCCAACATGCTGTTGCTTGTCTTACCTCATATGATCTCTTCAAAATGTG 2115  
 Db 681 GluArgGlyThrHisIleGlyLeuLeuAlaAsnProHisSerIleTyrSerGluMetTrp 700  
 QY 2116 CATACACAGACACCGGTGTCAGAACCATGATTAACCCCAATGGGAACAAAGAA 2175  
 Db 701 HisThrGlnSerSerArgValGlnAsnHisAspAsnProLysTrpGluAlaLysLysGlu 720  
 QY 2176 AATATATCAAGAGAGAGAAAGAAAGAACTACAGAAGAAATTTGCAATAGTGTAAA 2235  
 Db 721 AsnIleSerLysGluGluGluArgLysLysLeuGlnGluGlnIleValAlaSerValLys 740  
 QY 2236 GGCTGTGGAACCTGTTGCTGC 2256  
 Db 741 GlyCysGlyAsnCysSerCys 747

## RESULT 3

US-08-895-522-3  
 ; Sequence 3, Application US/08895522  
 ; Patent No. 5858719

## GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.  
 APPLICANT: Shah, Purvi  
 TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE  
 TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Dr.  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA

## ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible

## OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/895,522

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J  
 REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0336 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

## TELEX:

INFORMATION FOR SEQ ID NO: 3:

## SEQUENCE CHARACTERISTICS:

LENGTH: 694 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: Genbank

CLONE: 1167982

US-08-895-522-3

## Alignment Scores:

Pred. No.: 0 Length: 694  
 Score: 3342.00 Matches: 656

```

AF133659 (1-2345) X US-08-895-522-3 (1-694)
Percent Similarity: 97.41% Conservative: 20
Best Local Similarity: 94.52% Mismatches: 18
Query Match: 78.56% Indels: 0
DB: 2 Gaps: 0

```

|    |      |   |      |
|----|------|---|------|
| QY | 175  | GAACATTAAAGATATCACTGACAGAGATTGGGAAAAGCAATTCCAGACAGTTCTTA          | 234  |
| Db | 1    | Glusterleuathgahmtrhrthglnglnaistfipdilyvaspsphmsnergrglnleu      | 20   |
| QY | 235  | GATGCTGCAAAGGCTCTCCAGATATGGCCACTGATGAAAAAGAGCATGTTGGCATGT         | 294  |
| Db | 21   | AspIatThrlysalaleuGlntrhtrProleuileglulysarghrcystrph:sgily       | 40   |
| QY | 295  | CATGACGAGAGAGACTCCCAACAGACCACAAAGAGGTTAAAGAATGTGTACTCGG           | 354  |
| Db | 41   | HispIacglylglyleuHsthrAspPrologluclyleuLysaspValAspThrA:cg        | 60   |
| QY | 355  | AAATCATTAAGAACAAATGCTTTCTTAATGTGGCCCAAGACAGGCGCAATCTACAGCT        | 414  |
| Db | 61   | LysIlelleuLysAlaMetLeuSerTyrValTrpProglunbsprgrProbspLeuArgAla    | 80   |
| QY | 415  | AGAGTTGCATTTCGCTGGGATTTTGGGTGGTCAAGGCCATGAATTTGTGTTCC             | 474  |
| Db | 81   | ArgValAlaIleSerLeuenglyPheLeuGlyGlyAlaLysAlaMetAsnIleValPro       | 100  |
| QY | 475  | TTCAAGTTTAAATGTGTGTRGACAGGCTCAACAGATGTGGGAAACATGTGAAACCG          | 534  |
| Db | 101  | PheMetPheLysTyrAlaValAspSerLeuAsnGlnMetSerLysMetLeuAsnLeu         | 120  |
| QY | 535  | AGTGATGCACCAATATACAGTTGCACACCATGGCAACAGAGTTTGTGATGGCTATGTGA       | 594  |
| Db | 121  | SerAspIaProAsnThrValAlaThrMetAlaThrAlaValLeuIleGlyTyrGlyVal       | 140  |
| QY | 595  | TCAAGAGCTGGAGCTGCTTTTTTTAAAGAGTTTCAAAATGCAGTATTGGCAAGTATGCC       | 654  |
| Db | 141  | SerTrgAlaGlyAlaAlaPhePheAsnGlyValAlaArgAsnAlaValaPheGlyLysValaAla | 160  |
| QY | 655  | CAGATTCAATCCGAGAAATAGCCAAAAATGCTTTCTCCATTTTCAACACCTGCATCTG        | 714  |
| Db | 161  | GlnAsnSerIleArgAlaGlyIleAlaLysAsnValPheLeuHisLeuHisAsnLeuAspLeu   | 180  |
| QY | 715  | GGTTTTCACTGAGCAGACAGACGGGAGCTTATCTPAAGGCTATTGACAGAGAAACAGG        | 774  |
| Db | 181  | GlyPheHisLeuSerArgGlnThrGlyAlaLeuSerLysValaIleAspArgGlyThrArg     | 200  |
| QY | 775  | GGATTCAGTTTGGCCGAGAGGCTTGGATTTAAATCTTCCATTCACATGTTTGAAGG          | 834  |
| Db | 201  | GlyIleSerPheValLeuSerAlaLeuValaPheAsnLeuEubProlleValPheGlnMet     | 220  |
| QY | 835  | ATGCTTGTCAAGTGATTTTGTATATCAAAATGGCGATGCCAGTTTGTCTTGGTAAACCTT      | 894  |
| Db | 221  | MetLeuValSerSerValLeuTyrTyrLysCysGlyAlaGlnPheAlaLeuValThrIleu     | 240  |
| QY | 895  | GGAACACTTGGTACATACACAGCAATTCACAGTTGCAGTCAACAGGGTGGAGAACTAGATT     | 954  |
| Db | 241  | GlyThrLeuGlyAlaTyrTrhAlaPheThrValAlaValaThrArgTrpArgThrArgPhe     | 260  |
| QY | 955  | AGATTAAGAAATGAACAAGACAGATTAATGATGACAGGTAATCTGCTTATACCTCACTG       | 1014 |
| Db | 261  | ArgIleGlnMetAsnLysAlaAspAsnAspAlaGlyAsnAlaAlaIleAspSerLeuLeu      | 280  |
| QY | 1015 | AATTATGAACCTGTGAAGTATTTTAAATGAAGAATATGAGACACAGATATGATGGA          | 1074 |
| Db | 281  | AsnTyrGlnThrValLysTyrPheAsnAsnGlnLysTyrGlyAlaGlnArgTyrAspGly      | 300  |
| QY | 1075 | TTTTTGAAGACGATGAGACTGCTCATTTGAAAAGTACCTTACTGTGGCTATGTGCAGAC       | 1134 |
| Db | 301  | PheLeuLysThrTyrGlnTrhAlaSerLeuLysSerThrSerThrIleuAlaMetLeuAsn     | 320  |
| QY | 1135 | TTTGTGCAAGTGCTATTTTCAATGTGCTTAAACAGCTATAATGGTGTGCTGCCAGTCA        | 1194 |

|   |      |   |      |
|---|------|---|------|
| D | 321  | PheIyGlnAsnAlaIlePheSerValGlyLeuThrAlaIleMetValLeuAlaSerGln         | 340  |
| Q | 1195 | GAATTCGGCAGGTACCTTACTGTGGAGATCTAGTATGGTGAATGAGACTCTTTT              | 125  |
| D | 341  | GlyIleValAlaGlyAlaLeuThrValGlyAspLeuValMetValAsnGlyLeuLeuPhe        | 360  |
| Q | 1255 | CAGCTTCATTACCCCTGAACTTTCTGGGAACGTGTATATGAGAAGACTGACAGCACTC          | 1311 |
| D | 361  | GlnLeuSerLeuProLeuAsnPheLeuGlyThrValTyrArgGlnThrArgGlnAlaLeu        | 380  |
| Q | 1315 | ATACATATGAAACCCCTGTTTACTCTATACCAAGCTGACACCACCAATTTAAAGCAAGG         | 1374 |
| D | 381  | IleAspPheAsnThrLeuPheThrLeuLeuValAspThrArgIleLeuAspIlyVal           | 400  |
| Q | 1375 | ATGCACTCTCCCTTCAGATCACACCAACAGACACTACCGTGGCTTTGATATATGCAAT          | 1433 |
| D | 401  | MetIaIreProProLeuGlnIleThrProGlnThrAlaThrValAlaIaPheAspAsnValHis    | 420  |
| Q | 1435 | TTTGAAATACATTGAGGGCCGAAAAGTCCTTATGGAATATCCTTTGAAGTCCCTGCAGGA        | 1491 |
| D | 421  | PheGlnTyrIleGlnGlyGlnIlyValIleuAsnGlyValSerPheGlnIValIProAlaGly     | 440  |
| Q | 1495 | AAGAAATGGCCCAATTGAGAGAGGTACTGGGTGCAGGGAAAGACAACTAGTGGAGCATTA        | 1555 |
| D | 441  | LysIlyValAlaIleValGlyGlySerGlySerGlyIlySerThrIleValAlaIrgLeuLeu     | 460  |
| Q | 1555 | TTTCGCTTCATGAGCCTCAAAAGGGTAGCAATTATCTTGCTGGTCAAAATATACAGAT          | 1614 |
| D | 461  | PheArgPheTyrGlnProGlnIlySerIleTyrLeuAlaGlyGlnAsnLeuGlnAsp           | 480  |
| Q | 1615 | GTGAGCTGAAAAGCCTTGGAGGGCAGTGGAGTGTACTCAGAGATCTGTCTCTTC              | 1674 |
| D | 481  | ValSerLeuGlnSerLeuArgAlaValAlaGlyValIValProGlnAspAlaValLeuPhe       | 500  |
| Q | 1675 | CATTAATCTATTATATACACCTCTTATATGAAAATCATAGTCCTTCACTGAGGAAGG           | 1733 |
| D | 501  | HisAsnThrIleTyrIlyAsnLeuLeuTyrGlyAsnIleAsnAlaSerProGlnGlnAla        | 520  |
| Q | 1735 | TATCAGTGGCCAAATATAGCTGAGCTTCATGTAGTGCATCTTCGAAATGGCAGATGAT          | 1795 |
| D | 521  | TyrIaValAlaIlySerLeuAlaGlyLeuAlaIleAspAlaIleLeuArgMetProHisGlyTyr   | 540  |
| Q | 1795 | GACACCCAACTAGGGGAAACGAGGACTCAAGCTTTCAGAGGAGAAAAGCAAGTACGA           | 1855 |
| D | 541  | AspArgIleValGlyGlnArgGlyLeuIlySerGlyGlyIlySerGlnArgValAla           | 560  |
| Q | 1855 | ATTGCAAGACCATTTTGAAGGACCCCCAGTCATATCTTATATGATGAGACTTCTCATGG         | 1911 |
| D | 561  | IleIaIArgAlaIleLeuIlyAsnProProValIleLeuTyrAspGlnAlaThrSerSer        | 580  |
| Q | 1915 | TTTAATTCGATTCTGAAAGGATATTTCTTTGGGCCATGAAAGATGGTGTCAACACAGA          | 1974 |
| D | 581  | LeuAspSerIleThrGlnGlnIlyIleuGlyAlaMetArgAspValAlaIlyHisAspG         | 600  |
| Q | 1975 | ACTTCTATTTTTCATGACACAGATTTGTCAACAGTGGTGTGATGACATGAAATCATTTGTC       | 2033 |
| D | 601  | ThrSerIlePheIleIaHisArgLeuSerThrValValAspAlaAspIuIleIleVal          | 620  |
| Q | 2035 | TTGGATACAGGGTAAAGTACCGCAACGTGTACCCACCATGTTTGGTCTGTACCTCAT           | 2094 |
| D | 621  | LeuSerGlnGlyIlyValAlaGlnIaIArgGlyIlyThrHisTyrGlyLeuIleuAlaAsnSerSer | 640  |
| Q | 2095 | AGTATCATTCAGAAATGTGGCATACACAGACGCGGTGGCAGAACATATTAACCC              | 2155 |
| D | 641  | SerIleTyrSerGlnIleTyrPheIlyIleThrGlnIleAspAspValaGlnAsnGlnAspSerLeu | 660  |
| Q | 2155 | AAATGGGACCAAGAAAGAAATATATATCAAGAGAGAGAGAGAAAGAAATCAACAGA            | 2214 |
| D | 661  | GlyIlyThrAspAlaIlyIlySerGlnSerLeuSerIlySerGlnGlnGlnIlyIlySerGlnGln  | 680  |
| Q | 2215 | GAATTCCTCAATAGTGTGAAGGCTGTGGAACCTGTCTGTC                            | 2256 |
| D | 681  | GlnIleValaIleSerValIlySerGlyCysGlyAsnCysSerCys                      | 694  |

```

RESULT 4
US-09-195-391-3
; Sequence 3, Application US/09195391
; Patent No. 6080842
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
; TITLE OF INVENTION: TRANSPORT PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,391
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/895,522
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0336 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 694 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1167982
; US-09-195-391-3

Alignment Scores:
Pred. No.: 0 Length: 694
Score: 3342.00 Matches: 656
Percent Similarity: 97.41% Conservative: 20
Best Local Similarity: 94.52% Mismatches: 18
Query Match: 78.56% Indels: 0
Gaps: 0

AF133659 (1-2345) x US-09-195-391-3 (1-694)
QY 175 GAGTCATTAAAGATGATCATGAGAGATTGGAAAAAGGCATTCAGACAGTCTTCA 234
Db 1 GlnserleuArgAsnThrThriGlnGlnArgLysAspAsnSerArgGlnleu 20
QY 235 GATGCTGCAAGGCTCTCCAGGTATGCGCATGATGAAAAAGAGACATGTTGGCATGCT 294
Db 21 AspAlaThrLysAlaLeuGlnThrTrpProleuIleGlnLysArgThrCysTrpHisGly 40
QY 295 CATGAGAGAGAGACTCCACAGACAGCCAAAGAGGTTAAAGATGTTGATCTCGG 354
Db 41 HisAlaGlyGlyGlyLeuHisThrAspProLysGlnGlyLeuLysAspAlaAspThrArg 60
QY 355 AAAATCATAAAGCAATGCTTCTTATGTTGTCGCCAAAGACAGCCAGATCTAGAGCT 414

```

```

Db 61 LysIleIleLysAlaMetLeuSerThrValTrpProGlnAspArgProAspLeuArgAla 80
QY 415 AGAATTCGCAATTTGGCTGGGATTTTGGGTGGTCGCAAGGCAGATATATTGGTCCC 474
Db 81 ArgValAlaIleSerLeuGlyPheLeuGlyGlyAlaLysAlaMetAsnIleValValPro 100
QY 475 TTCATGTTTAAATATGCTGTAGACAGCTCAACAGATGTCGGGAAACATGCTGAACCTG 534
Db 101 PheMetThrLysThrAlaValAspSerLeuAsnGlnMetSerGlyAsnMetLeuAsnLeu 120
QY 535 AGTATGACCAACCAATATACATTGCAACCATGGCAACAGACTTGTATGGCTATGGTGA 594
Db 121 SerAspAlaProAsnThrValAlaThrMetAlaThrAlaValLeuIleGlyThrGlyVal 140
QY 595 TCAAGACTGAGCTGCTTTTAAAGAGTTGGAATGAGATTTGGCAAGTGGCC 654
Db 141 SerArgAlaGlyAlaAlaPheAsnGlnValArgAsnAlaValPheGlyLysValAla 160
QY 655 CAGAAATCAATCCGAGAAATAGCAAAATGCTCTTCCATCTTCACAACTGATCTG 714
Db 161 GlnAsnSerIleArgAlaGlyIleAlaLysAsnValPheLeuHisLeuHisAsnLeuAspLeu 180
QY 715 GGTTCACCTGACGACAGACAGACGGAGCTTATCTAAAGCTATTGACAGAGAACAGG 774
Db 181 GlyPheHisLeuSerArgGlnThrGlyAlaLeuSerLysAlaIleAspArgGlyThrArg 200
QY 775 GGTATCAGTTTGTCCGAGAGCTTGTGATTTTAAATGCGTCCCATCATCTGGAAGTG 834
Db 201 GlyIleSerPheValLeuSerAlaLeuValPheAsnLeuLeuProIleValPheGlnMet 220
QY 835 ATGCTTGTCAGTGTGTTTGTATTAACAATGCGGTGCCAGTTGCTTGTGAACCTT 894
Db 221 MetLeuValSerSerValLeuThrLysThrLysGlyAlaGlnPheAlaLeuValThrLeu 240
QY 895 GGAACACTGTGTACATACACAGCATTCACAGTTCAGTCACAGCGTGAGAACATTAATT 954
Db 241 GlyThrLeuGlyAlaLysThrAlaPheThrValAlaValThrArgTrpArgThrArgPhe 260
QY 955 AGAATAGAAATGAACCAAGCAGATTAATGATGACAGTAAAGTGTATGACTCATGCTG 1014
Db 261 ArgIleGlnMetLeuLysAlaAspAsnAspAlaGlyAsnAlaAlaIleAspSerLeuLeu 280
QY 1015 AATTATGAACTGTGAGATATTTAATGAAGAATATGAAAGACAGACAGATATATGGA 1074
Db 281 AsnThrGlnThrValLysThrPheAsnAsnGlnLysThrGlnAlaGlnArgThrAspGly 300
QY 1075 TTTTGAAGACGATGAGACTGCTTCATTGAAAGTACTTACTGCTGCTAGCTGAAC 1134
Db 301 PheLeuLysThrThrGlnThrAlaSerLeuLysSerThrSerThrLeuAlaMetLeuAsn 320
QY 1135 TTTGCTCAAAATGCTATTTTCAAGTGGCTTAAACAGCTAATAGTGTCGCGAGTCAG 1194
Db 321 PheGlyGlnAlaAlaIlePheSerValGlyLeuThrAlaIleMetValLeuAlaSerGln 340
QY 1195 GGAATGTGGAGAGTACCTTACTGTTGAGATCTAGTAATGTAATGAGATGCACTGCTTTT 1254
Db 341 GlyIleValAlaGlyAlaLeuThrValGlyAspLeuValMetValAsnGlyLeuLeuPhe 360
QY 1255 CAGCTTTCATTAACCTCGAATCTTCTGGGAACTGTATATAGAGAGACTAGACAAGCTC 1314
Db 361 GlnLeuSerLeuProLeuAsnPheLeuGlyThrValLysArgGlnThrArgGlnAlaLeu 380
QY 1315 ATAGATATGAACACCTTGTACTCTACTCAAGGTATACACCCCAATTAAGACAAAGTG 1374
Db 381 IleAspMetAsnThrLeuPheThrLeuLeuLysValAspThrArgIleLysAspLysVal 400
QY 1375 ATGCAATCTCCCTTCAGATCACACACAGACAGTACCGTGCGCTTGAATATGTCAT 1434
Db 401 MetAlaProProLeuGlnIleThrProGlnThrAlaThrValAlaPheAsnValHis 420
QY 1435 TTTGAATACATTTGAGGGGACAAAGTCTTATGTAATATCTTTGAAGTCCCTGACGA 1494

```

Db 421 PheGluTrpIleGluGlnGlnValLeuAsnGlnValIleSerPheGluValProIleGly 440  
QY 1495 AAGAAAGTGGCCATTGAGAGAGTAGGGGTGAGGAAAGACAAATAGTAGAGCTATTA 1554  
Db 441 LysIleValAlaIleValIleGlySerGlySerGlySerGlySerThrIleValArgLeuLeu 460  
QY 1555 TTTCGGCTTATGAGCCTCAAAAGGTAGCATTTATCTGTGGTCAAAATATCAAGAT 1614  
Db 461 PheArgPheTrpGluProGlnLysGlySerIleTrpLeuAlaGlyGlnAsnLeuGlnAsp 480  
QY 1615 GTGAGCTGGAAAGCCTTCGAGAGGAGTAGGGTAGCTGACCTGAGATGGCTCTCTC 1674  
Db 481 ValSerLeuGlnSerLeuAlaArgAlaValAlaGlyValAlaProGlnAspAlaValLeuPhe 500  
QY 1675 CATATACTATTATTAACAACCTTTATATGAGAAACATCAAGTGGTCACTGAGAAATG 1734  
Db 501 HisAsnThrIleTrpTrpAsnLeuLeuTrpGlyAsnIleAsnAlaSerProGlnGlnVal 520  
QY 1735 TATGCAAGTGGCAAAATTAAGTGGACCTTCATGATGCAATTTCTGCAATGGCAATGAT 1794  
Db 521 TyrIleValAlaLysLeuAlaGlyLeuHisAspAlaIleLeuArgMetProHisGlyTyr 540  
QY 1795 GACACCCAGTAGGGGGAACGAGCACTCAAGCTTCAGAGAGGAAAGCAAAAGTAGTGA 1854  
Db 541 AspThrGlnValGlyGlnArgGlyLeuLysLeuSerGlyGlyGlnLysGlnArgValAla 560  
QY 1855 ATTGCAAGAGCATTGTAAGAGACCCCACTCATCTATGATGTAAGTAGTCACTTCG 1914  
Db 561 IleLeuArgAlaIleLeuLysAsnProProValIleLeuTrpArgGlnAlaThrSerSer 580  
QY 1915 TTAGATTGCAATTACTGAAGAGACTATTCTTGTCGCATGAAGATGGTGCACACAGA 1974  
Db 581 LeuAspSerIleThrGlnGlnThrIleLeuGlyValMetArgAspValIleLysHisArg 600  
QY 1975 ACTTCTATTGTCATGACACAGATTGTCACAGTGGTGTATGAGATGAAATGATGTC 2034  
Db 601 ThrSerIlePheIleAlaHisArgLeuSerThrValAlaAspAlaAspGlnIleLeuVal 620  
QY 2035 TTGGATCAGGGTAAAGTAGGCGAAGCTGTGTAACCACTGCTTGTGCTAAACCTCAT 2094  
Db 621 LeuSerGlnGlnLysValAlaGlnArgGlyThrHisIleTrpGlyLeuLeuAlaAsnSerSer 640  
QY 2095 AGTATCTATTCAGAAATGTGGCATATCAAGACAGCCGTGTGCAGAACATGATTAACCC 2154  
Db 641 SerIleTrpSerGlnMetTrpHisThrGlnSerAsnArgValGlnAsnGlnAspSerLeu 660  
QY 2155 AATGGAACCAAGAAAGAAATATATCCAAAGAGAGGAAAGAAAGAACTTCAAGAA 2214  
Db 661 GlyTrpAspAlaLysLysGlnSerLeuSerIleGlnGlnArgLysLysLeuGlnGln 680  
QY 2215 GAAATTGTCAATAGTGTGAAGAGCTGTGAACACTGTCGTGC 2256  
Db 681 GlnIleValAsnSerValLysGlyCysGlyAsnCysSerCys 694

RESULT 5  
US-08-895-522-4  
; Sequence 4, Application US/08895522  
; Patent No. 5858719  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Shah, Purvi  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE  
; TITLE OF INVENTION: TRANSPORT PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/895,522  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0336 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 694 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 575393  
US-08-895-522-4  
Alignment Scores:  
Pred. No.: 4,17e-147 Length: 694  
Score: 1516.50 Matches: 304  
Percent Similarity: 67.83% Conservative: 122  
Best Local Similarity: 48.41% Mismatches: 169  
Query Match: 35.65% Indels: 33  
DB: 2 Gaps: 7  
AF133659 (1-2345) x US-08-895-522-4 (1-694)  
QY 355 AATAATATAAACAATGCTTTCTTATGTGTGGCCCAAGACAGCCAGATCTACAGCT 414  
Db 94 LysIleLeuLysAspLeuPheArgTyrIleTrpProLysGlyAsnAsnLysValArgIle 113  
QY 415 AGAGTTGGCATTTCGGTGGATTTTGGGTGGTGAAGGCCATGATATGTGGTCCC 474  
Db 114 ArgValLeuIleAlaLeuGlnLeuIleSerAlaLysIleLeuAsnValGlnValPro 133  
QY 475 TTGATGTTTAAATATGCTGTAGACAGCTCAACAGATGTGGGAAACATGCTGAACCTG 534  
Db 134 PhePhePheLysGlnThrIleAspSer-----MetAsnIle 145  
QY 535 AGTGAGCACCAATATCAGTTGCAACCATGGCAACA-----GCAAGTTCGATTGGCTAT 588  
Db 146 AlaTrpAspAspProThrValAlaLeuProAlaAlaIleGlyLeuThrIleLeuCysTyr 165  
QY 589 CGGTATCAAGAGCTGAGCTGCTTTTAAAGAGTTCGAAATCGAATTTGGCAAG 648  
Db 166 GlyValAlaArgPheGlySerValLeuPheGlyGlnLeuArgAsnAlaValPheAlaLys 185  
QY 649 GTAGCCAGAAATTCAGATCGAAGATAGCCAAATATGCTTTCTTCATCTTCAACACTG 708  
Db 186 ValAlaGlnAsnAlaIleArgThrValSerLeuGlnThrPheGlnHisLeuMetLysLeu 205  
QY 709 GATTCGGGTTTCACTGACGACAGACAGACGGGAGCTTTATCTAAGCTATTTGACAGAGA 768  
Db 206 AspLeuGlyTrpHisLeuSerArgGlnThrGlyGlyLeuThrArgAlaMetAspArgGly 225  
QY 769 ACAAGGGTATGATGTTTGGTCCGAGTGGCTTGGTATTTAATCTCTCCATCATGTTT 828  
Db 226 ThrLysGlyIleSerGlnValLeuIleAsnValPheHisIleIleProIleSerPhe 245  
QY 829 GAAGTATGCTTGTCAAGTGGTGTGTTGTATTCAAATGGCGTCCAGATTGCTTGGTA 888

```

Db 246 GlutSerValIcylGlyLeuThrTrpGlnPheGlyAlaSerPheAlaIle 265
Qy 889 ACCCTGGACACCTGGTACATACACAGCTTACAGTTCAGTACACGGTGAAT 948
Db 266 ThrPheSerThrMetLeuLeuTrpSerIlePheThrIleValThrAlaTrpThr 285
Qy 949 AGATTGAAATGAGATGACCAACAGATATGATGAGTAAATGCTGTATGATCA 1008
Db 286 HisPheArgAspAlaAsnLysAlaAspAsnLysAlaAlaSerValAlaLeuAspSer 305
Qy 1009 CTGCTGAATTATGAAACCTGTGAAGTATTTAATGAAAGATATGAAACAGATAT 1068
Db 306 LeuIleAsnPheGluAlaValLysTrpPheAsnAsnGluLysTyrLeuAlaAspLysTrp 325
Qy 1069 GATGATTTTGAACAGCTATGAGACTGCTTATGAAAGTACCTTACTCTGCTATG 1128
Db 326 AsnGlySerLeuMetAsnLysTrpArgAspSerGlnIleLysValSerGlnSerLeuAlaPhe 345
Qy 1129 CTGAACCTTGGTCAAGTCTATTTTCAGTCTGGTTTAAAGCTATATGCTGCTCGCC 1188
Db 346 LeuAsnSerGlyGlnAsnLeuIlePheThrAlaLeuThrAlaMetCysTrpGly 365
Qy 1189 AGTCAGGAAATGTGCAAGTACCTTACTGTGAGATCTAGTATGATGATGAGACTG 1248
Db 366 CysThrGlyValIleGlyGlyAsnLeuThrValGlyAspLeuValLeuIleAsnGlnLeu 385
Qy 1249 CTTTTCAGCTTTCATTAACCTGTAACCTTTCGGAACTGTATATAGAGACTGACAA 1308
Db 386 ValPheGlnLeuSerValProLeuAsnPheLeuGlySerValTrpArgAspLeuLysGln 405
Qy 1309 GCATCCATGATATGACACCTGTTTACTTACTACAGTGAAGTGAACCAATTAAGAC 1368
Db 406 SerLeuIleAspMetGlnThrLeuPheLysLeuArgLysAsnGluValLysIleLysAsn 425
Qy 1369 AAAGTATGATGATCTCCCTTCAGATCAACCAAGACAGCT-----ACCGTGGCTTT 1422
Db 426 -----AlaGluArgProLeuMetLeu---ProGluAsnValProTrpAspIleThrPhe 442
Qy 1423 GATATGTGATTTTGAATACATTGAGGGCCAGAAAGTCTTAGTGAATATCCTTTGAA 1482
Db 443 GluAsnValIleThrPheGlyTrpIleAspThrAspArgLysIleLeuLysAsnAlaSerPheThr 462
Qy 1483 GTCCCTGAGGAGAAAGAGTGGCCATTTGAGAGGAGTGAAGGAGGAGGAGGAGCAATA 1542
Db 463 IleProIleGlyTrpLysThrAlaIleValGlySerSerGlySerGlyLysSerThrIle 482
Qy 1543 GTAGGCTATTAATTCGCTTCTATGAGCCTCAAAAGGATGACATTATCTGCTGCA 1602
Db 483 LeuLysLeuValPheArgPheTrpAspProGlnSerGlyArgIleLeuIleAsnGlyArg 502
Qy 1603 AATATCAAGATGTAGCCTGGAAGCCTTCGAGGGCAGTGGAGTGTACTACAGAT 1662
Db 503 AspIleLysGluTrpAspIleAspAlaLeuArgLysValIleGlyValValProGlnAsp 522
Qy 1663 GGTGCTCTCTTCATTAATCTATTATTAACAACCTTATGAGAACTCAAGCTTCA 1722
Db 523 ThrProLeuPheAsnAspThrIleTrpGluAsnValLysPheGlyArgIleAspAlaThr 542
Qy 1723 CTTGAGGAAGTATGACGTGCAAAATTAAGCTGACATTCATGATGCAATTTCTGCAAT 1782
Db 543 AspGluGluValIleThrValAlaGluLysAlaGlnLeuAlaProLeuIleLysLysLeu 562
Qy 1783 CCACATGATATGACCAACCAAGTGGGAGAGAGACTCAAGCTTTCAGAGAGAGAAAG 1842
Db 563 ProGlnGlyPheAspThrIleValGlyGluArgLysLeuMetIleSerGlyGlyLys 582
Qy 1843 CAAAGATGACATTCGACAGAGCAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1902
Db 583 GlnArgLeuAlaIleAlaArgValLeuLeuLysAsnAlaArgIleMetCysPheAspGln 602
Qy 1903 GCTACTTATCGTTAGATTCGATTAACGAAAGACTATTTCTTGCCATGAGAGAT-- 1959

```

```

Db 603 AlaThrSerAlaLeuAspThrHisThrGluGlnAlaLeuLeuArgThrIleArgAspAsn 622
Qy 1960 ---GTGGCAAAAGACAGAACTTCTATTCTTCTTTCACACAGATTTGTCACAGTGGTGA 2016
Db 623 PheThrSerGlySerArgGlnSerValTrpIleAlaHisArgLeuArgThrIleAlaAsp 642
Qy 2017 GCAGATGAATCATTTGTTGGATCAGGGTAAAGGTACCCGAACTGTATCCACCATGCT 2076
Db 643 AlaAspLysIleIleValLeuAspAsnGlyArgValArgGluGluLysHisLeuGlu 662
Qy 2077 TTGCTGTATACCTCATAGTATCTATTCAGAAATGTGGCATACAGACAGACCGCTGTG 2136
Db 663 LeuLeuAlaMetProGlySerLeuTrpArgGluLeuTrpThrIleGln----- 678
Qy 2137 CAGAACATGATTAACCCCAATGGAGCAAGAAAGAAATATATATCCAAAGAGAGGAA 2196
Db 679 -----GluAspLeuAspHisLeuGlnAsn 686
Qy 2197 AGAAAGAACTACAGAAATTT 2220
Db 687 GluLeuLysAspGlnGlnGluLeu 694

RESULT 6
US-09-195-391-4
: Sequence 4, Application US/09195391
: Patent No. 6080842
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Shah, Purni
: APPLICANT: Corley, Neil C.
: TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
: TITLE OF INVENTION: TRANSPORT PROTEIN
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/195,391
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/895,522
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0336 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 694 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: GenBank
: CLONE: 575393
: US-09-195-391-4

Alignment Scores: 4.17e-147 Length: 694
Pred. No.:

```

Score: 1516.50 Matches: 304  
Percent Similarity: 67.83% Conservative: 122  
Best Local Similarity: 48.41% Mismatches: 169  
Query Match: 35.65% Indels: 33  
DB: 7 Gaps: 3

AF133659 (1-2345) x US-09-195-391-4 (1-694)

QY 355 AAAAAATCAAAAGCAATGCTTTCTTATGTGTGGCCAAAGACAGCCAGATCTACAGACT 414  
Db 94 LysilleuysaspheupheAIGTyrIletrProlysgIysaenulysValArgIle 113  
QY 415 AGAGTGGCATTTGGCTGGGATTTTGGGGTGGCCAAAGCCATGAATATTCGCTTCC 474  
Db 114 ArgValIleuIlealaleuGlyLeuIleuIleSerAlaIlyIleuAenValGlnValPro 133  
QY 475 TTCACTTTAAATATGCTGTAGACAGCTCAACAGATGTCCGGAACATGCTGAACCTG 534  
Db 134 PhephephepysgInthrIleApsSer-----MetAsnIle 145  
QY 535 AGTATGACCAAAATACAGTTGCAACCATGACACA-----GCAGTTCTGATTCGCTAT 588  
Db 146 AlatrPaspaspProthrValAlaLeuProAlaAlaIleGlyLeuThrIleLeuGlyTyr 165  
QY 589 GGTGTATCAAGAGCTGAGCTGCTTTTAAAGAGTTGAAATGCAATTTGGSCAG 648  
Db 166 GlyValAlaIleArgPheGlySerValIleuPheGlyGluIleuArgAsnAlaValPheAlaIlys 185  
QY 649 GTAGCCCAAGATTCATCCGAAGAATAGCCAAATGCTTTCCTCCATCTTCACAACTG 708  
Db 186 ValAlaGlnAsnAlaIleArgThrValSerLeuGlnThrPheGlnIleuMetLysLeu 205  
QY 709 GATCTGGGTTTTCCTGACAGACAGACAGCGGAGCTTTATCTAAAGCTTATTCAGACAGA 768  
Db 206 AspLeuGlyIleThrIlePheSerArgGlnThrGlyLeuThrArgAlaMetAspArgGly 225  
QY 769 ACAAGGGATACAGTTTGTCTGAGTGGCTTGGTATTAATCTTCCCATCANGTTT 828  
Db 226 ThrIleGlyIleSerGlnValIleuThrAlaMetValPheHisIleIleProIleSerPhe 245  
QY 829 GAATGATGCTGTCACTGAGTGTGTTTGTATTAACAATGCCGTCGCTTGGTGA 888  
Db 246 GluIleSerValValCysGlyIleLeuThrIleGlnPheGlyAlaSerPheAlaAlaIle 265  
QY 889 ACCCTTGAACACTTGTATACACAGACTTCACAGTTCAGTCAGACAGCGTGGAGACT 948  
Db 266 ThrPheSerThrMetLeuLeuTyrSerIlePheThrIleGlyThrAlaIlePheArgThr 285  
QY 949 AGATTTGAATAGAAATGAACAAGCATATGATGATGATGATGATGATGATGATGAT 1008  
Db 286 HisPheArgIleArgAspAlaLeuLysAlaAspAsnLysAlaIleSerValAlaLeuAspSer 305  
QY 1009 CTCTGGAATATATGAACCTGGAAGTATTTATTAATATGAAGAATGACACAGATAT 1068  
Db 306 LeuIleAsnIleGlnAlaValLysTyrPheAsnAsnGlnLysTyrLeuAlaAspLysTyr 325  
QY 1069 GATGATTTTGAAGACGATGAGACTGCTTCATTTGAAAAGTCACTCTACCTGCGCATG 1128  
Db 326 AsnGlySerIleuMetLeuTyrArgAspSerGlnIleLysValSerGlnSerIleuAlaPhe 345  
QY 1129 CTGAACCTTGGTCAAACTGCTATTTTCAGTGTGGTTAAACAGTATATGATGCTGCC 1188  
Db 346 LeuAsnSerGlyGlnAsnLeuIlePheThrThrAlaLeuThrAlaMetMetTyrMetGly 365  
QY 1189 AGTCAGGAATTTGGGAGTACCTTACTGTTGGAGATCTAGTATGGAATGGACTG 1248  
Db 366 CysThrGlyValIleGlyGlnAsnLeuThrValGlyAspLeuValIleuIleAsnGlnLeu 385  
QY 1249 CTTTTCAGCTTTCATTAACCTGCACTTCTGGGAACTGTATATGAGAGCATAGACA 1308  
Db 386 ValPheGlnLeuSerValProLeuAsnIleuGlySerValTyrArgAspLeuLysGln 405  
QY 1309 GCATCATATGATATGAACACTTGTCTTACTCAAGGTAGACCCCAATTAAGAC 1368

Db 406 SerLeuIleAspMetGluThrLeuPheLysLeuArgLysAsnGlnValLysIleLysAsn 425  
QY 1369 AAGATGATGCATCTCCCTTCAGATCACACGACAGACT-----ACCGTGGCCTTT 1422  
Db 426 -----AlaGluArgProIleuMetLeu---ProGlnAsnValProTyrAspIleThrPhe 442  
QY 1423 GATATATGTCATTTGAATATCATTTGAGGGCCAGAAAGTCCCTTATGTAATATCTTTGAA 1482  
Db 443 GluAsnValThrPheGlyTyrHisProAspArgLysIleLeuLysAsnIleSerPheThr 462  
QY 1483 GTCCCTTCAGGAAAGAAAGTGGCATTGTRAGAGGTAGTGGGTGACGAGGAAACACAAATA 1542  
Db 463 IleProAlaGlyTyrPysThrAlaIleValGlySerSerGlySerGlyLysSerThrIle 482  
QY 1543 GTAGGCTATTTATTTGGCTTCTATGAGCCCTCAAAAGGTAGCTTTATCTGTGTCGCA 1602  
Db 483 LeuLysLeuValPheArgPheTyrAspProGlnSerGlyArgIleLeuIleAsnGlyArg 502  
QY 1603 AATATACAGATGTGAGCCCTGGAAGCCCTTCGAGGGCAGTGGAGTGTACTCAGAT 1662  
Db 503 AspIleLysGluTyrAspIleAspAlaLeuArgLysValIleGlyValValProGlnAsp 522  
QY 1663 GCTGTCTCTTCATTAATCTATTTATTAACAACCTTTATATGAGAAACATCAGTGTCTCA 1722  
Db 523 ThrProLeuPheAsnAspThrIleTyrGlnAsnValLysPheGlyArgIleAspAlaThr 542  
QY 1723 CCTGAGGAAGTGTATCCAGTGGCAAAATAGCTGAGCTTCATGATCCATTTCTCGAATG 1782  
Db 543 AspGlnGluValIleThrValValGluLysAlaGlnLeuAlaProLeuIleLysLysLeu 562  
QY 1783 CCAGATGATATGACACCAAGTAGGGAAGACGAGACTCAAGCTTCAGAGAGGAAAG 1842  
Db 563 ProGlnGlyPheAspThrIleValGlyGluArgGlyLeuMetIleSerGlyGlyLys 582  
QY 1843 CAAGAGTAGCAATTCAGAGAGCATTTTGAAGACCCCTCAGTCACTATGATGAA 1902  
Db 583 GlnArgLeuAlaIleAlaArgValIleuLysAsnAlaArgIleMetPhePheAspGln 602  
QY 1903 GCTACTTCACTGATATTCGATTTACTGAAAGACATTTCTGTCGCCATGAAGAT--- 1959  
Db 603 AlaThrSerAlaLeuAspThrHisThrGlnAlaLeuAlaArgThrIleArgAspAsn 622  
QY 1960 ---GTGTCAAACACAGACTTCTATTTTCATTTGACACAGATGTCACAGGTGTGAT 2016  
Db 623 PheThrSerGlySerArgThrSerValTyrIleAlaHisArgLeuArgThrIleAlaAsp 642  
QY 2017 GCAGATGAATATCATTTGTCTGATCAGGGTAGAGTACCGAAGCTGTGATCCCACTAGGT 2076  
Db 643 AlaAspLysIleIleValLeuAspAsnGlyArgValArgGlnGluLysHisLeuGln 662  
QY 2077 TTGCTGTCAACCTCATATGATCTATTTCAAGAAATGTGGCATATACAGACAGCCGTG 2136  
Db 663 LeuLeuAlaMetProLysSerLeuTyrArgGluLeuTyrThrIleGln----- 678  
QY 2137 CAGAACCATGATTAACCCCAATGGGAAAGCAAGAAATATATATCAAAAGAGAGAA 2196  
Db 679 -----GluAspLeuAspHisLeuGlnAsn 686  
QY 2197 AGAAGAAACTACAGAGAAATT 2220  
Db 687 GluLeuLysAspGlnGlnGlnLeu 694  
RESULT 7  
US-09-212-247C-4  
; Sequence 4, Application US/09212247C  
; Patent No. 6391603  
; GENERAL INFORMATION:  
; APPLICANT: POMPEYUS, Markus; SUEHLBERGER, Harald; JOSEFFKEN, Hans  
; and GARCIA, DOVAL, Jose Luis Revuelta; JIMENEZ, Alberto;  
; and GARCIA, Maria Angeles Santos  
; TITLE OF INVENTION: Genes of purine biosynthesis from Ashbya Gossypii  
; and the use thereof in microbial riboflavin

```

;
;      synthesis
;      NUMBER OF SEQUENCES: 21
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Keil & Weinkauff
;      STREET: 1101 Connecticut Avenue
;      CITY: Washington
;      STATE: D.C.
;      COUNTRY: USA
;      ZIP: 20036
;
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
;      COMPUTER: IBM AT-compatible, Pentium processor
;      OPERATING SYSTEM: Windows 95
;      SOFTWARE: WordPerfect version 6.1
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/09/212,247C
;      FILING DATE: 16-Dec-1998
;      CLASSIFICATION: <unknown>
;      INFORMATION FOR SEQ ID NO: 4:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 475 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 4:
;
US-09-212-247C-4
;
Alignment Scores:
Pred. NO.:      8.45e-115      Length:      475
Score:          1201.50        Matches:      234
Percent Similarity: 71.74%    Conservative: 96
Best Local Similarity: 50.87%  Mismatches:  127
Query Match:    28.24%        Indels:       3
DB:             4              Gaps:         2

AF133659 (1-2345) x US-09-212-247C-4 (1-475)
Oy 757 ATTGACAGAGCAAGGGGATGATGAGTTTGTCTGAGTGGTTGTAATTAATCTTCT 816
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MetaphaeglyCyslyseglylleseertyvalleuserAlaMetValPhehisIleIle 20
Oy 817 CCCATCATGTTTGAAGTGAAGTGTCTGTCAAGTGTGTTTGTATTAACAAATGCCGTCAG 876
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 21 ProlethrPhegluIleaserMetValCysglyIleleuthrTyGlnPhegllylaser 40
Oy 877 TTTCGTTGTTGACCTTGGAACCTTGATACATACAGCATTCACAGTTCACTGACATCA 936
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 41 PheAlaAlaIlethrPheSerThrMetleuLeuTySerIlePheThrPheArgThrThr 60
Oy 937 CGGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 996
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 AlaThrPheGlnThrArgPheArgPheArgPheArgPheArgPheArgPheArgPheArg 80
Oy 997 GCTATAGACTGCTGCTGATTAATGAAGCTGGAAGTATTAATGAAGTATGAAGTATGA 1056
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 81 AlaIleuAspSerleuIleAsnPheGlnAlaValTySerPheAsnAsnGlnulTySerleu 100
Oy 1057 GCACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1116
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 101 AlaAspLysTyHisIleThrSerleuMetLysTyArgAspSerGlnIleLysValSerGln 120
Oy 1117 ACTCGGCGATGCTGATGCTGCAAGTGTATTTTTCAGTGGTGGTTACAGCTATA 1176
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 SerleuAlaPheleuAsnThrIleGlnAsnLleuIlePheThrAlaIleuThrAlaIle 140
Oy 1177 ATGTGCTCGCAGTACAGGATTTGGACGATGACCTTACTGTTGAGATCTAGTAATG 1236
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 141 MetTyRmetAlaCysAsnGlyValMetGlnGlySerleuThrValGlyAspIleuValleu 160
Oy 1237 GTGAATGAGCTGCTTTTTCAGCTTTTCATTACCCCTGGAACCTTTTGGGAACTGTATATGA 1296
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 161 IleAsnGlnIleuValPheGlnIleuSerValProleuAsnPheleuGlySerValTyArg 180

```

```

Oy 1297 GAGACTAGACAGACACTCATATGATATGACACCTTGTTACTCTACTCAAGTAGACACC 1356
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 AspLeuLysGlnSerleuIleAspMetGlnSerleuPheLysGlnLysAsnGlnVal 200
Oy 1357 CAATTTAAAGCAAAAGTAGGACTCTCCCTTCAAGTCAACACAGACAGCTACCGTG 1416
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 201 ThrIleLysAsnSerProAsnAlaGlnAsnLeuProleu--HisLysProleuAspIle 219
Oy 1417 GCCTTGATATGTCATTTTGAATCATTTGAGGGCCAGAAAGCTTACGTAATATCC 1476
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 220 ArgPheGlnAsnValThrPheGlyTyAspProGlnArgArgIleLeuAsnAsnValSer 239
Oy 1477 TTGAGAGCTCGCAGGAGAAAGAGTGGCCATTGTAGACGACTAGTGGTCAAGGAAAGC 1536
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 240 PheThrIleProAlaGlyMetLysTyThrAlaIleValGlyProSerGlySerGlyLysSer 259
Oy 1537 ACAATAGTAGAGCTATTTATTTGCTTCTATAGACCTCAAAAAGGTACATTATCTTGCT 1596
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 260 ThrIleLeuLysLeuValPheArgPheTyGlnProGlnGlnGlyArgIleLeuValGly 279
Oy 1597 GGTCAAAATATACAGATGTGAGCGCTGAAAGCTTCGAGAGGCGAGTGGAGTGAATCT 1656
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 280 GlyThrAspIleArgAspLeuAspLeuSerleuArgLysAlaIleGlyValValPro 299
Oy 1657 CAGATGCTGTCTCTTCATTAATACTATTATTAACAACCTTTATATGGAACATCAGT 1716
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 300 GlnAspThrProleuPheAsnAspThrIleTyrGlnAsnValLysPheGlyAsnIleSer 319
Oy 1717 GCTTCACCTGAGGAAGTGTATGACAGTGGCAAAATTAAGTGGACTTCATGATGCAATCTT 1776
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 320 SerSerAspAspGlnIleLeuArgAlaIleGlnLysAlaGlnLeuThrLysLeuLeuGln 339
Oy 1777 CGAATGCCACATGATATGACACCCCAAGTAGGGGAGAGAGACTCAAGCTTCAGAGAGA 1836
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 340 AsnLeuProLysGlyAlaSerThrValValGlyLysGlyLeuMetIleSerGly 359
Oy 1837 GAAAGCAAGATGAGCAATTTGCAAGACCATTTTGAAGACCCCGCATCATCTAT 1896
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 360 GlnLysGlnArgLeuAlaIleAlaArgValleuLysAspAlaProleuMetPhePhe 379
Oy 1897 GATGAGCTACTTATGTTGATTCATTAATCTGAAGACATATCTTGGTCCATGAG 1956
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 380 AspGlnAlaThrSerAlaLeuAspThrHisThrGlnGlnAlaLeuHisIleGln 399
Oy 1957 GAT-----GTGTCAAACACAGAACTTCTATTTCATTGACACAGATGTCAACAGTG 2010
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 400 GlnAsnPheSerSerAsnSerLysThrSerValTyValAlaHisArgLeuArgThrIle 419
Oy 2011 GTTGATGAGATGAAATCATTTGCTTGATGATGAGGATGAGGATGAGGATGAGGATGAG 2070
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 420 AlaAspAlaAspLysIleIleValleuGlnGlnGlySerValArgGlnGlnGlyThrHis 439
Oy 2071 CATGTTGCTTGGTAAACCTTCATAGTATCTATTCAGAAATGTGGCATACACAGACG 2130
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 440 SerSerleuLeuAlaSerGlnGlySerleuTyArgGlyLeuThrAspIleGlnGlnAsn 459

RESULT 8
US-09-134-001C-4730
; Sequence 4730, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4730

```



```

; LENGTH: 582
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4730

Alignment Scores:
Pred. No.: 1.07e-62 Length: 582
Score: 695.50 Matches: 186
Percent Similarity: 51.41% Conservative: 123
Best Local Similarity: 30.95% Mismatches: 261
Query Match: 16.35% Indels: 31
DB: Gaps: 12

AF133659 (1-2345) x US-09-134-001C-4730 (1-582)

QY 352 CGGAAATCATAAAGCAATGCTTTCTATGTGTGCCCAAGACAGGCGCATCTACGA 411
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 ArgArgnctleuysArgTyLeuLysPheValysPro-----TyrArg 17
QY 412 GCTAGAGTGGCCATTTGCGTGGGATTTTGGGTGCGCAAG--GCCATGAATATTGTG 468
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 18 TyrArglellealatrille---lleValGlyllelleLysPheGlylleProMetleu 36
QY 469 GTTCCCTTACGTTAAATATGCTGTAGACAGCTCAACGATGCGGAAACATGCTG 528
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 37 lleProleuLeuileuTyTyAlaileasp-----GlyValile 49
QY 529 AACCTGAGTATGACCAATATAGCTTGCACATGCGAACAGAGCTTGTGATTGGC--- 585
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 50 AsnAsnHisserleuthrAsnGlnGlyLysPheSerHisleuGlyValAlaileGlyle 69
QY 586 -----TATGTGTATCAAGAGCTGAGCTGCTTTTAAAGAACTTGAAT 633
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 70 AlaleupheillePheleulleValAlargProleuileGluPhe-----lleArgGln 86
QY 634 GCATATATTGGCAGAGTACCAGAAATTCATCCGAAATAGCCAAATATGCTTCTC 693
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 87 TyreuleaGlntrPthrSerAsnLysilleLeuTyAspilleArgLysGlnLeuTyAsn 106
QY 694 CATCTCAACAACCTGATGCTGGTTCACCTGAGACAGACAGACAGGGA---GCTTATCT 750
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 107 HisleuGlnAlaLeuSerValArgPheTyAlaAsnAsnGlnValGlyGlnValIleSer 126
QY 751 AAGCTATGACAGAGGAGACAGGGGATACAGTTTGTCTGATGCTTGTGATTTAAT 810
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 127 ArgValIleAsnAspValGlnGlnThrLysAspHeilleuThrGlyLeuMetAsnIle 146
QY 811 CTTCCTCCCATCATGTTTGAAGTATCTGTACGTGGTGTGTTTGTATACAAATGCGGT 870
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 147 TrpLeuAspCysIleThrIlellelleAlaLeuSerIleMetPhePheLeuAspValLys 166
QY 871 GCCAGATTTGCTTGTGTAACCTTGAACACTTGTATACATACAGAGA---TTCACAGTT 927
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 167 LeuthrPheAlaAlaIlePheIlePheProPheTyrlleuThrValTyPhePhePhe 186
QY 928 GCATCTCACGCTGAGACACTAGATTAGATAGAAATGAACAAAGACAGATATGATGCA 987
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 GlyArgLeuArgLysLeuThrArgValArg-----serGlnAlaLeuAlaGluVal 203
QY 988 GGTATGCTGCTATAGACTACCTGCTGAATTAAGAACTGTAAGATTTTAAATATGAA 1047
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 204 GlnGlyPheLeuHisGlnArgValGlnGlyMetSerValIleLysSerPheAlaileGlu 223
QY 1048 AGATATGACAGACAGATATGATGATTTTGAAGACGATAGATGCTTCACTTGAAA 1107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 224 AspAsnGlnIleAlaLysAsnPheAspAsnHisAsnLysAsnPheLysGlnArgAlaPheGln 243
QY 1108 AGTACCTTACTGCTGCTATGCTGAATCTTGTGTCAAAGTCTATTTTCAAGTGGGTTTA 1167
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 244 HisThrArgTrpAsnAlaLysPheAlaAlaIleAsnThrValThrAspLeuGlyPro 263
QY 1168 ACAAGCTTAATGGTCTGCCGACGAGGAAATGTCGACAGTACCTTACTGTTGGAGAT 1227
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 264 llelleValIleGlyValGlySerTyLeuAlaIleThrGlySerIleThrValGlyThr 283
QY 1228 CTAGTAATGGTGAATGACTGCTTTTTCAGCTTTCATTAACCCCTGAACCTTCTGGAACT 1287
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 LeuAlaAlaPheValGlyTyLeuGlnGlnLeuPheGlyProleuArgAlaGluValSer 303
QY 1288 GTATATAGAGAGACTAGACAGACTCATATGATATGAACCTTGTTTACTTACTCAAG 1347
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 304 SerPheThrThrLeuThrGlnSerPheAlaSerMetAspArgValPheGlnLeuMetAsp 323
QY 1348 GTAGACACCCAAATTAAGCAAGATGATGCACTCCCTCCCTGAGATCAACACACAGACA 1407
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 324 GluAspTyAspIleLysAsnGlylleGlyAlaGlnProIleLysIleSer-----Lys 341
QY 1408 GCTACCGTGGCTTTGATATATGTCATTTTGAATATACATTGAGGCGCAGAAA---GTCTT 1464
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 342 GlyGlnIleAspLeuLysHisValSerPheLysTyTrAsnGlnAsnGlyLysGluValLeu 361
QY 1465 AGTGAATATCTTTGAAAGTCCCTGACAGAAAGAAAGTGGCCATTTGAGAGGTAGTGGG 1524
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 362 HisAspIleAsnLeuThrIleAsnLysGlyGluThrValAlaPheValGlyMetSerGly 381
QY 1525 TCAGGAAAAGCACATATAGTGAAGGCTATATTCGCTTCTATAGACCTCAAAAAGGTAGC 1584
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 382 GlyGlyLysSerThrleuIleAsnleuIleProArgPheTyTrAspValThrGlnGlyGlu 401
QY 1585 ATTATCTTGTCTGCTCAAAATATACAAAGATGTAGCCTGGAAGCCTTGGAGGGGAGTG 1644
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 402 lleleuIleAspHisHisAsnValLysAspPheLeuThrGlySerleuAsnGlnIle 421
QY 1645 GAGGTGATCTGAGATGCTGCTCTTTCATTAATATATTATTAACAACCTTATAT 1704
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 422 GlyLeuValGlnGlnAspAsnIleleuPheSerAspThrValLysGlnAsnIleleuLeu 441
QY 1705 GGAATCATGCTCTTCACTGAGAGAGTATAGCAGTGGCAAAATTAGTGCATTTCT 1764
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 442 GlyArgProAspAlaThrAspAspGluValAlaGlnAlaLysMetAlaAsnAlaHis 461
QY 1765 GATGCAATCTTTCGATATGCGACATGATATGACACCCAGTAGGAGACAGACTCAAG 1824
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 462 AspPheIleSerAsnLeuProAsnGlyTyAspThrGluValGlyGlnArgGlyValLys 481
QY 1825 CTTCAGAGAGGAGAAAGACAAAGATGACATTTGCAAGGCCATTTTGAAGACCCCCA 1884
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 482 LeuSerGlyGlyGlnLysGlnArgLeuSerIleAlaArgIlePheLeuAsnAsnPropio 501
QY 1885 GTATATCTATATGATGACTTATCTGTTAGATTGATTCGATTAAGTGAAGACTATTTCT 1944
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 502 ValLeuIleLeuAspGlnAlaThrSerAlaLeuAspLeuGluSerGlnAlaIleIleGln 521
QY 1945 GGTGCCATGAGAGATGGTGCACAAACAGAACTTCTATTTTTCATTCACACAGATGTCA 2004
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 522 GluAlaLeuAspValLeuSerLysAspArgGlnThrLeuIleValAlaHisAspGluSer 541
QY 2005 ACAGTGTGATGACAGATGAATATGCTTGTGATCAGGATGAAGGTAGCCGAGCTGTG 2064
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 542 ThrIleThrHisAlaAspArgIleValAlaMetGluAsnGlyArgIleValGluThrGly 561
QY 2065 ACCCACCAGTGTGCTGCTAAACCTCATATGATATTCAGAAATGCGGACATACAG 2124
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 562 ThrHisGlnGlnleuIle---AsnLysArgGlyAlaTyGluHisleuTySerIleGln 580
QY 2125 AGC 2127
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 581 Asn 581

```

```

RESULT 9
US-08-772-270A-4
; Sequence 4, Application US/08772270A
; Patent No. 6019984
; GENERAL INFORMATION:
; APPLICANT: MacInnes, Janet
; APPLICANT: Ricciatti, Paul

```



```
Db 630 AsplGluAlaThrSerAlaLeuAspArgLysSerGlnHisIleIleMetArgAsnMetHis 649
Qy 1957 GATGTGTCACACAGACTTCTATTTCATGTGCACACAGATGTGTAACAGTGGTGTAT 2016
Db 650 GlnIleCysArgGlyArgThrValIleIleIleAlaHisArgLysSerThrValLysAsn 669
Qy 2017 GCAGATGAATCATGTGCTTGATCAGGGTAGAGTACGCCAGACGTGTATCCACCATGT 2076
Db 670 AlaAspArgIleIleValMetGlnLysGlyGlnIleValGlnGlnGlyLysHisGln 689
Qy 2077 TTGCTTGCTAACCTCATAGTATCTATTTCAGAAATGTGGCTAACAGAC 2127
Db 690 LeuLeuAlaAspProAsnGlyLeuThrIleSerGlnLeuHisGlnLeuGlnSer 706

RESULT 10
US-09-134-001C-5611
; Sequence 5611, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5611
; LENGTH: 580
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5611

Alignment Scores:
Pred. No.: 9,86e-57 Length: 580
Score: 637.50 Matches: 174
Percent Similarity: 50.52% Conservative: 118
Best Local Similarity: 30.10% Mismatches: 239
Query Match: 14.99% Indels: 47
DB: 4 Gaps: 12

AF133659 (1-2345) x US-09-134-001C-5611 (1-580)
Qy 406 CTACGACCTAGAGTGGCATTGCTGGAGATTTTGGGTGGCAAGGCCATGATATT 465
Db 27 TLeAlaAlaValIleIleThrSerLeuGlySerLeuSerGly-----Leu 41
Qy 466 GTGGTCC-----TTCATGTTTAATAATGCTGTAGACAGCTTCAC--- 507
Db 42 LeuValProLeuPheThrGlyArgLeuValAspLysPheSerValSerSerIleAsnTrp 61
Qy 508 -----CAGATGTCGGGAACATGCTGAACCTGAGTATGACCAATAATACAGTTGA 558
Db 62 GlyMetIleAlaIlePheGlySerIlePhe----- 71
Qy 559 ACCATGGCAACAGCATTTGATTTGGCTATGCTGATTCAGAGCTGAGCTGCTTTT 618
Db 72 ---LeuValAsnAlaLeuLeuSerGlyIleGlyLeu----- 82
Qy 619 ACGAAGTTGGAATGAGATATTGGCAAGGTACCCAGAAATTCACCAAGAAATAGCC 678
Db 83 -----TyrLeuLeuSerLysIleGlyGlnLysIleIleTyrAlaIleArg 97
Qy 679 AAAAATCTTTCTCATCTTTCACAACTGGATCTGGATTTTTCACCTGACAGACAGAG 738
Db 98 SerLeuLeuTrpGlnHisIleIleGlnLeuLysMetProPhePheAspLysAsnGlnSer 117
Qy 739 GGAAGCTTATCTAAAGCTATTGACAGAGAAAGGGGATACGTTTGTCTGAGTGCT 798
```

```
Db 118 GlyGlnLeuMetSerArgLeuThrAspPheThrLysValIleAsnGlnPheIleSerGln 137
Qy 799 TTGGATTTTATCTTCTCCCATCATGTTTGAAGTACGCTTGTCAGTGGTGTATTGAT 858
Db 138 LysLeuProAsnLeuLeuProSerValLeu---ThrLeuIleGlySerLeuValMetLeu 156
Qy 859 TACAAATGCGGTGCCAGTTTGTGTTGAACCTTGGAAACCTTGTATACATACACAGCA 918
Db 157 PheIleMetAspTrpLysLeuThrLeuThrPheIleThrIleProValPhe----- 174
Qy 919 TTCACATTCGACATGCACACGCTGAGAACCTAGATTAGAAATAGAACCAAGCAGAT 978
Db 175 IleLeuIleIleValProLeuGlyArgValMetGlnLysIleSerThrAsnThrGlnSer 194
Qy 979 AATGATGACGTAAATGCTGTATA-----GACTCAGCTGTAATTATGAACGTGAG 1032
Db 195 GlnIleAlaAsnPheSerGlyLeuLeuGlyArgValLeuThrGlnMetArgLeuValLys 214
Qy 1033 TATTTTAATGAATGAAGATATGAAGACAGAGATATGATGATTTTGAACGTATGAG 1092
Db 215 ValSerAsnThrGlnLysGlnLeuAspAsnAlaHisThrAsnLeuLysIleTyr 234
Qy 1093 ACTGCTTCATGAAAGTACTACTGCTGCTATGCTGTAACCTTGGTCAAAAGTCTATT 1152
Db 235 ArgLeuGlyLeuLysGlnAlaLysIleSerAlaValGlnProIleSerGlyValVal 254
Qy 1153 TTGAGTGTCCGTTTAAACAGCTATATGCTGCTGCCAGTCAGGGAATTGTGGCAGTACC 1212
Db 255 MetLeuLeuThrIleAlaIleIleLeuGlyPheGlyAlaLeuGlnIleGlyThrGlyAla 274
Qy 1213 CTTACTGTGGAGATCTAGTATGATGTAATGAGCTGCTTTTTCAGCTTTCATTAACCTTG 1272
Db 275 IleThrProGlyThrLeuIleAlaMetIlePheTyrValIleGlnLeuSerMetProLeu 294
Qy 1273 AACTTCTGGAACTGTATATAGAGACTAGACAAAGCACTCAATGATATGAACACTTG 1332
Db 295 IleAsnLeuSerThrLeuValIleThrAspTyrLysAlaValGlyAlaSerSerArgIle 314
Qy 1333 TTACTCTACTCAAGGTAGACACCAAAATTAAGACAAAGTATGACATCTCCCTTCAG 1392
Db 315 TyrGlnIleMetGlnGlnProIleGlu---ProThrGlnAlaLeuSerGlnSerLysAsp 333
Qy 1393 ATGACACACACAGACACTACCGTGGCTTGTATGATGTGCAATTTTGAATCATTTGAGGCG 1452
Db 334 ValThrIleIleAspGlyGlnLeuValPheGlnHisValAspPheLysTyr---AspVal 352
Qy 1453 CAGAAAGTCTTAGTGAATATGCTTTGAAGTCCCTGCGAGAAAGAAAGTGGCCATTGA 1512
Db 353 LysLysIleLeuGlnLysAspValSerPheSerIleProGlnGlyGlnValSerAlaPheVal 372
Qy 1513 GGAAGTAGTGGGTCAAGGAAACCAATATGAGGCTATTTTGGCTTCATGAGCCT 1572
Db 373 GlyProSerIleSerGlyLysSerThrIlePheAsnLeuIleGlnLysMetTyrAspIle 392
Qy 1573 CAAAAGGTAGCATTTATCTTGTGTCGTAATAATATACAGATGTGAGCCTGGAAGCCTT 1632
Db 393 GlnArgGlyAspIleLysTyrGlnAsnGlnSerIlePheAspIleProLeuSerLysTrp 412
Qy 1633 CGAAGGACAGTGGAGTGAACCTCAGAGTGTCTCTCTTCATATATATCTATTATTAC 1692
Db 413 ArgThrLysIleGlyTyrValMetGlnSerAsnSerMetMetSerGlyThrIleAspAsp 432
Qy 1693 AACCTCTTATATGGA---AACATCAGTGTCTACCTGAGGAAGTATGACAGTGGCAAA 1749
Db 433 AsnIleLeuTyrGlyIleAsnArgLysValAspAspGlnLysLeuIleGlnTyrAlaLys 452
Qy 1750 TTAGCTGACCTTCATGATGCAATTTCTGCAATGCCACATGATATATGACCAAGTAGGG 1809
Db 453 LeuAlaAsnCysHisAspPheIleMetGlnPheAspGlnGlyTyrThrAspThrMetValGly 472
Qy 1810 GAACGAGACTCAAGCTTTCAGAGAGAAAGCAAGAGTACCAATTTGACAGAGCCATT 1869
Db 473 GlnArgGlyLeuLysLeuSerGlyGlnArgGlnArgIleAspIleAlaArgSerPhe 492
```

QY 1870 TTGAAGGACCCCGCTACTCTATGTAGAGCTTACGTTAGATTCGATCTACT 1929  
Db 493 VallysaenProAspIleuLeuLeuLeuSpGluAlaThrAlaAsnLeuAspSerGluSer 512  
QY 1930 GAAGAGCTATTCTTGGTCCATGGAAGATGTGTCAAACACAGAACTTCAATTTTCATT 1989  
Db 513 GluLeuLysIleGlnGluAlaLeuGluThrLeuMetGluGlyArgThrThrValValIle 532  
QY 1990 GCACACAAATTTCTCAACAGCTGTTATGCATGAAATCATCTTGTGATAGGGTAG 2049  
Db 533 AlaHisAspGluSerThrIleLysLysAlaGlyGlnIleValPheIleAspLysGlyGlu 552  
QY 2050 GTAGCCGAAGCTGTACCCACCATGTTTGTCTTGAACCCCTCATGATCTAT 2103  
Db 553 ValThrGlyLysGlyThrHisIleGluLeuMetAlaSer--HisAspLysIleYr 569

RESULT 11  
US-08-772-270A-12  
Sequence 12, Application US/08772270A  
Patent No. 6019984  
GENERAL INFORMATION:  
APPLICANT: MacInnes, Janet  
APPLICANT: Ricciardi, Paul  
APPLICANT: Mallard, Bonnie  
APPLICANT: Rosendal, Soren  
TITLE OF INVENTION: NOVEL BACTERIAL PREPARATIONS, METHOD FOR  
TITLE OF INVENTION: PRODUCING SAME, AND THEIR USE AS VACCINES  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bereskin & Parr  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/772,270A  
FILING DATE: December 23, 1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Michelle  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 6580-81  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 711 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Actinobacillus pleuropneumoniae  
US-08-772-270A-12

Alignment Scores:  
Pred. No.: 1e-54 Length: 711  
Score: 618.50 Matches: 165  
Percent Similarity: 51.78% Conservative: 126  
Best Local Similarity: 29.36% Mismatches: 242  
Query Match: 14.54% Indels: 29  
DB: 3 Gaps: 10  
AF133659 (1-2345) x US-08-772-270A-12 (1-711)

QY 463 ATGTGGTTCCCTTCATGTTTAAATATGCTGTAGACAGCCCTCAACAGATGTCGGA--- 519  
Db 173 LeuIleThrProLeuPhePheGlnValValMetAspLysValLeuValHisArgGlyPhe 192  
QY 520 AACATGCTGAACCTGAGTATGACACCAATACAGTTGACACCATGCGAACAGCGATTCTG 579  
Db 193 SerThrLeuAsn-----ValIleThrValAlaLeuSerValVal 205  
QY 580 ATTGCGTATGCTGATCAAGAGCTGAGCTCTTTTAAACGAAGTTCGAATGACGATA 639  
Db 206 ValIlePheGluIleValLeuSerGly-----LeuArgThrTyrIle 219  
QY 640 TTGGCAAGTAGCCCAAAATTCATCCAGAAATAGCCAAAATGCTTTCTTCATCTT 699  
Db 220 PheSerHisSerThrSerArgIleAspValGluLeuGlyAlaLysLeuPheArgHisIleu 239  
QY 700 CACAACTGATCTGGGTTTCCACTGAGCAGACAGACGGAGCTTATCTAAGGCTATT 759  
Db 240 LeuAlaLeuProIleSerTyrPheGluAsnArgValGlyAspThrValAlaArgVal 259  
QY 760 GACAGAGGAACAAGGGGATACGTTTGTCTGTAGTGCTTGTATTTAATCTTCCC 819  
Db 260 ArgGluLeuAspGlnIleArgAsnPheLeuThrGlyAlaAlaLeuThrSerValLeuAsp 279  
QY 820 ATCATGTTTGAAGTAGATGCTGTGAGTGCTTTTGTATTAACAATGGCGCCAGTTT 879  
Db 280 LeuLeuPheSerPheIlePhePheAlaValMetIrrpYrYr-----SerProLysLeu 297  
QY 880 GCTTGGTACCTTGGACACTTGTGATACACAGCATTCACAGTTGCAGTGCACAGCG 939  
Db 298 ThrIleValIleLeuLeuSerLeuProCysTyrIleAlaIrrSerIlePheIleSerPro 317  
QY 940 TGGAGAACTAGATTGTAAGAAATGAACAACAGATAT--GATGACGATATGCT 996  
Db 318 IleLeuArgArg--ArgLeuAspGluLysPheAlaArgAsnAlaAspAsnGlnSerPhe 336  
QY 997 GCTATAGACTCAGCTGCAATTATGAAACTGTGAAGTATTTAAATGAAAGATATGAA 1056  
Db 337 LeuValGluSerValSerAlaIleAspThrIleLysAlaLeuAlaValThrProGlnMet 356  
QY 1057 GCACAGAGATATGATGCAATTTTGAAGCGTATGAGACTGCTTATGAAAAGTACTCT 1116  
Db 357 ThrAsnIleTrrAspLysGlnLeuAlaSerTyrValSerAlaAspPheArgValThr-- 375  
QY 1117 ACTCGGCTATCGCACTTGTGCTGCAAGTCTATTTTCACTGCGGTTTAACGACTA 1176  
Db 376 -----ValLeuAlaThrIleGlyGlnGlnValGlnLeuIleGlnLysThrValMet 393  
QY 1177 ATGGTG-----CTGCCAGTCAAGGAATTTGTGCGAGGTACCTTACTGTGGAGAT 1227  
Db 394 IleIleAsnLeuTrrPheGlyAlaHisIleLeuValIleSerGlyAspLeuSerIleGlyGln 413  
QY 1228 CTAGTATGATGATGACGCTGCTTTTCACTTTCATTAATCCCTGAACTTTGGGAAT 1287  
Db 414 LeuIleThrPheAsnMetLeuSerGlyGlnValIleAlaProValAlaArgLeuAlaGln 433  
QY 1288 GTATATAGAGAGACTAGACAGCACTCATGATATGAAACACTGTTTACTCTCTCAAG 1347  
Db 434 LeuTrrGlnAspPheGlnGlnValGlyIleSerIleThrArgLeuGlyAspValLeuAsn 453  
QY 1348 GTAGACACCCAAATTTAAGACAAAGTGTGACTCTCCCTTCAGATCACACACAGACA 1407  
Db 454 SerProThrGlu-----AsnTyrGlnGlnLysLeuSerLeuProGlnIlePhe 469  
QY 1408 GCTACCGTGCGCTTGTATATGTCATTTGAATAC--ATTGAGGCCAGAAAGTCTT 1464  
Db 470 GlyAspIleAlaPheLysHisIleArgPheArgTrrLysProAspAlaProIleIleLeu 489  
QY 1465 AGTGAATATCTTGAAGTCCCTGCAGAGAAAGAGTGGCATTTGAGAGGTAGTGGG 1524  
Db 490 AspAspValaLeuLeuSerValLysGlnGlyValIleGlyIleValGlyArgSerGly 509  
QY 1525 TCAGGAAAACACAAATAGTAGGCTATTATTTGCTTATGAGCCTCAAAAGGTTAGC 1584

```
|||||
Db 510 SerGlyVsSerThrLeuThrLysLeuLeuGlnArgPheTyrlleProGlnAsnGlyGln 529
1585 ATTATCTTGCTGCAATATATACAGATGATGAGCTCGAAGCCCTTGAGGGAGCTG 1644
:::
Db 530 ValLeuIleArgGlyHisAspIleuAlaLeuAlaAspProAsnThrPheLeuArgGlnIle 549
1645 GAGAGTGATCTCAGAGATGCTGCTCTTCATATATATCTATTTATTCACACCTCTTAT 1704
GlyValValLeuGlnAspAsnValLeuLeuAsnArgSerIleArgAspAsnIleAlaLeu 569
1705 GGAACATCATGCTCTCACTCGAGAGAGTGTATGACAGTGGCAAAATTAGCTGACTTAT 1764
:::
Db 570 ThrAspProSerMetSerMetGluArgValIleTyrlalAlaLysLeuAlaGlyAlaHis 589
1765 GATCAATTTCTTCAGATGCCACATGATATGACACCCCAAGTAGGGGAACGAGACTCAAG 1824
AspPheIleSerGluLeuArgGluGlyTyrlAsnThrIleValGlyGluLeuGlyAlaGly 609
1825 CTTTCAGAGAGAGAAAGCAAGATAGCAATTCGACAGCCATTTTGAGAGACCCCA 1884
LeuSerGlyGlyGlnArgGlnArgIleAlaIleAlaArgAlaLeuValAsnAsnProArg 629
1885 CTCATACCTCTGATGAAGCTACTTATCATCGTTAGATTCTGAAAGACTATTTCTT 1944
:::
Db 630 IleuIlePheAspGluAlaThrSerAlaLeuAspTyrlGlnSerGlnIleIleMet 649
1945 GGTGCAGTAAAGATGTGGTCAAAACAGACTCTATTTTCATTCGACAGATTCGCA 2004
GlnAsnMetGlnLysIleCysHisGlyArgThrValIleIleIleAlaHisAspLeuSer 669
2005 ACATGTGTTGATGAGATGAATCATGTCTTGGATCAGGTAGGTAGGTAGGTAGGT 2064
ThrValLysAsnAlaAspArgGlyIleValMetGluLysGlyHisIleValGlnGlnGly 689
2065 ACCACCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2124
LysHisAsnGlnLeuLeuGlnAsnGlnAsnGlyLeuTyrlTyrlLeuAsnGlnLeuGln 709
2125 AGCAGC 2130
Db 710 SerAsn 711

RESULT 12
: Sequence 2, Application US/08394880B
: Patent No. 5705352
: GENERAL INFORMATION:
: APPLICANT: Peery, Robert B.
: APPLICANT: Skatrud, Paul L.
: TITLE OF INVENTION: Multiple Drug Resistance Gene Of
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company/Patent Division
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: US
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/394,880B
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant G., Thomas
: REGISTRATION NUMBER: 35784
: REFERENCE/DOCKET NUMBER: X-9682
```

```
TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317) 276-2459
: TELEFAX: (317) 277-1917
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 791 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-394-880B-2

Alignment Scores:
Pred. No.: 3.13e-52 Length: 791
Score: 594.50 Matches: 227
Percent Similarity: 42.84% Conservative: 129
Best Local Similarity: 27.32% Mismatches: 302
Query Match: 13.98% Indels: 173
DB: 1 Gaps: 32

AF133659 (1-2345) x US-08-394-880B-2 (1-791)
QY 17 TGCAATCTTGCGCGTGGCGCGCGCGCG----- 46
Db 33 CysValIleSerAlaArgHisSerAlaArgAsnGlyLeuIleArgGlnPheSerGlyCys 52
47 -----CTGCTTCGAAAGCGCCGCGACCTCCGATTCGTATCCGCTTTAGTCTCTG 100
Db 53 SerGlySerIleSerAsnSerCysAsnProArgProTyrlArgSerAlaIleThrSerLeu 72
101 TTACGGCTCAGTCCGACGTGAGGC----- 127
Db 73 LeuSerAlaAsnValCysSerLysGlyValSerAlaValGlnProArgPheLeuSerThr 92
128 ---CAATCAACTCGGCGCTTGGAACCGCTGAGCT----- 163
93 ValArgLeuPheSerThrSerGlnArgSerLeuGlnProLysSerAsnValLysSerThr 112
164 -----ACGAGATTCAGAGTCATTAAAGATATCATGCGACATGCGAGATTGGGAA 211
Db 113 GlyGlyGlnValValArgProGluLeuHisGlnAspGlnIleHis---GluAspIleGlu 131
212 AAGCAATTGACAGACAGTCTTGAATGCTGCAAAGCTCTCCAGATGAGCCACTGATAG 271
Db 132 LysGlyPheGluLeuSerGluArg-AlaAlaGlnAlaAlaGlnVal-----AsnLeu 149
272 AAAAGAGACATGTGGCATGTGCATGTCAGAGAGAGACCTCCACAGACCCAAAAGAG 331
149 erAlaLysLeuAlaLysAspGlyAlaAlaGly-----LysLysAlaG 163
332 GGTAAAGATGTGATCTCGGAAATCATAAAGCAATGCTTTCTTATGTGTGGCCCA 391
Db 163 LysPheLysGluIle---TrpArgLeuLeuLeuIleAla----- 174
392 AAGACAGCCAGATCTACAGACTAGAGTTCATTCGCTGAGATTTTGGTGTGGCAA 451
175 -----ArgProGluAlaLys---LysLeuAlaLeuAlaPheLeuPheLeuValSer 192
452 AGGCCATGAATATGTTGTTCCCTTCATGTTTAAATATGCTGTAGACAGCCTCAACAGA 511
192 erGlyIleThrMetSerIleProPheSerIleGlyLysIleMetAspThrSerThrLys 511
512 TGTGCG-----GGAAACATGCTGAACCTGAGTATGACCAAAATACAGTTGCAACCA 562
Db 212 LathrThrArgGlyGlyAsnGlnLeuPheGlyLeuSerLeuProMetPheTyrlGlyAla 232
563 TGCCAAACAGCAGTTCTGATGCTATGCTGTATCAAGAGCTGAGAGCTGCTTTTATACG 622
232 euAla-----GlyIleLeuThrLeuGlyAlaAla---AlaAsn 244
QY 623 AAGTTGAAATGACGATTTGGCAGAGTACCCAGAAATTAATCCGAAGATAGCCAAA 682
244 yrlValArgIleIleIleLeuArgIleValGlyGluArgIleValAlaArgLeuArgSer 264
```

|    |      |  |  |       |      |
|----|------|--|--|-------|------|
| Qy | 683  | ANGTCCTTCATCTTCACAAACCTGGATCGGCTTTT                          |  | ----- | 720  |
| Db | 264  | yleuPheargIlnthrPheValGlnsnpIaGluPhePheaspIaAsnArgValGlyA    |  | ----- | 284  |
| Qy | 721  | -----CACCTGACGACAGACGGAGCTTATCTTAAGCTATTGACAGA----           |  | ----- | 765  |
| Db | 284  | spleuIleSerArgIeuserserSeraphrIleIleValGlyysSerIleThrGlnAmL  |  | ----- | 304  |
| Qy | 766  | -----GGACMACGGGT-----A                                       |  | ----- | 778  |
| Db | 304  | euserAspGlyLeuArgIaIaValSerGlyIaIaValaglyPheGlyLeuMetAlaTyrv |  | ----- | 324  |
| Qy | 779  | TCAGTTTTGCTCGAGCTCTGGATTTTATCTCTTCCACATGTTTGAAGAGAGC         |  | ----- | 838  |
| Db | 324  | alSerIeuLysIeuserserIleLeuAlaIeuLeuPro-----ProI              |  | ----- | 339  |
| Qy | 839  | TGTGACGTGCTGTTTGTATTAC--AAATCGCGTCCAGTTTGGTATGACCTTG         |  | ----- | 895  |
| Db | 339  | IeglyIeuGlyIaPhePheTyrglyArgIaIaIeArgAsnLeuserArgGlnIleGlnA  |  | ----- | 359  |
| Qy | 896  | GAACTGTGTACATACACACGATTCACAGTTGCATCACGCGTGAAGAACTAATTTA      |  | ----- | 955  |
| Db | 359  | rgAsnLeuglyThyrLeuThr-----                                   |  | ----- | 365  |
| Qy | 956  | GAATGAATGAATGAACACACATATATATGACGTTATGCTGCTATAGACCTACGCTGA    |  | ----- | 1015 |
| Db | 366  | -----LysIleAlaGlnGluArgLeuGlyA                               |  | ----- | 374  |
| Qy | 1016 | ATTATGAACCTGTGAAGTATTTATATATGAAGAATGTGAACGACAGATATGATGAT     |  | ----- | 1075 |
| Db | 374  | snValLysThrSerIlnserPheAlaGlyIuValLeuGluValArgArgyrrAsnAsnG  |  | ----- | 394  |
| Qy | 1076 | TTTTG--AAGACGTATGACCTGCTTCAITGAAAGTAACTCTTACTTGCTGATGCTGA    |  | ----- | 1132 |
| Db | 394  | InvalArgIysIlePheGlu-----LeuGlyLysGlnSerLeuIleSerAlaT        |  | ----- | 411  |
| Qy | 1133 | ACTTGTGTCAAGTCTATTTTCAGTGTGCGTTTAAAGCTATATAGTGGTGGCAGTGC     |  | ----- | 1192 |
| Db | 411  | hPhePheSerSerThrGlyPheAlaGlyAsnMetThrIleLeuAlaIeuLeuTyrvAlG  |  | ----- | 431  |
| Qy | 1193 | AGGGA-----ATTGTGCGACGTACCTTACTGTTGAGATCTA-----               |  | ----- | 1230 |
| Db | 431  | IyGlyGlyMetValGlnSerGlyAlaIleThrIleGlyIuLeuThrSerPheLeuMet   |  | ----- | 451  |
| Qy | 1231 | -----GTAATGTGTGAATGAGACTCTTTCAGCTTTCATTATCCCTGAACTTTGGGA     |  | ----- | 1285 |
| Db | 451  | yrThrAlaTyralaGlySerSerMetPheGlyLeuSer-----S                 |  | ----- | 464  |
| Qy | 1286 | CTGTATATATGAGAGCTAGACACAGCACTCATAGTATGAAACACTTGTTTACTCATGCA  |  | ----- | 1345 |
| Db | 464  | ePheTyrserIleuLeuMetLysGlyValGlyIaIaAserArgLeuPheGluLeuGlnA  |  | ----- | 484  |
| Qy | 1346 | AGGTAGACACCCCAATTAAAGCAAAAGTANGGATCTCCCTT-----CAGATCA        |  | ----- | 1396 |
| Db | 484  | sPArgGlnProThrIle-----SerProThrLysGlyIuLysValA               |  | ----- | 498  |
| Qy | 1397 | CACCAACAGACATACCGTGCCTTTGATATGTGCATTTTGAATAC-----ATTGAGG     |  | ----- | 1456 |
| Db | 498  | IaSerAlaArgIyProIleArgPheGluAsnValThrPheSerTyrrProThrArgProA |  | ----- | 518  |
| Qy | 1451 | GCCAGAAAGCTCTTAGTGGATATCCCTTTGAAGCTCGTGAAGAAAGAAAGTGGCATTG   |  | ----- | 1510 |
| Db | 518  | IaValProIlePheArgAspLeuAsnPheGluIleProGlnGlyThrAsnValAlaIleV |  | ----- | 538  |
| Qy | 1511 | TAGAGATGTAGTGGGACAGGAAAGACAAATAGTGAAGCATATATTGCTTCTTATGAGC   |  | ----- | 1570 |
| Db | 538  | AlGlyProSerGlyIuLysSerThrIleAlaSerIleLeuLeuArgPheTyrserP     |  | ----- | 558  |
| Qy | 1571 | CTCAAAAGGTATGACTTATCTTGCTGTGCTAAAATATACAAGATGAGACTGGAAGCC    |  | ----- | 1630 |
| Db | 558  | roThrGlnGlyArgValLeuIleGlyLysAspIleThrIaIaMetAsnAlaLysSerL   |  | ----- | 578  |
| Qy | 1631 | TTTCGAGGGAGCTGGAGTGTACTCAGAGATGCTGTCTTCCTTCCATTAATCTATTATT   |  | ----- | 1690 |









```
OY 1309 GCACATAGATATGACACCTGTTTACTCTACTC-----AAGGTAGACACC 1356
Db 364 -----IlehelysilelleAspAnlyseProserlleAspSer 376
OY 1357 CAAATTAAGACAAAGTATGCGATCTCCCTTCAGATCACACACAGACTACCGTG 1416
Db 377 TyserLysSerGlyHslyseProAspAnlyleLys-----GlyAsnLeu 391
OY 1417 GCCTTGATTAATGTCATTTTGAATACATTTAGAGGCAG-----AAGTCTTAGTGA 1470
Db 392 GluPheArgAsnValHisPheSerTyProserArgLysGluValLysIleLeuLysGly 411
OY 1471 ATATCCTTTGAAGTCCCTGCAGAGAAAGAGTGCATTTGTAGAGAGTAGTGGTCAGAG 1530
Db 412 LeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGly 431
OY 1531 AAAAGCACATATGTAGAGCTATTATTGGCTTATAGAGCTCAAAAAGGTAGCATTTAT 1590
Db 432 LysSerThrThrValGlnLeuMetGlnArgLeuTyAspProThrGlnGlyMetValSer 451
OY 1591 CTTCCTGGTCAAAATATATCAAGATGAGAGCTGAAAAGCTTCGAGAGGCGAGTGGAGTG 1650
Db 452 ValAspGlyGlnAspIleArgThrIleAsnValArgPheLeuArgGluIleIleGlyVal 471
OY 1651 GTACCTCAGAGATGCTGTCTCTTCATAATATCTATTATTACACCTCTTATATGSAAC 1710
Db 472 ValSerGlnGluProValLeuPheAlaThrThrIleAlaIleValAsnIleArgTyGlyArg 491
OY 1711 ATCAGTCTTCACCTGAGAGAAATGTATGAGTGGCAAAATTAGCTGACTTCATGATGCA 1770
Db 492 GluAsnValThrMetAspGlnIleGlnLysAlaValLysGluAlaAsnAlaTyAspPhe 511
OY 1771 ATCTTCGAATGCCACATGATGATGACCCAAAGTAGGGGAGACGAGACTCAAGCTTCA 1830
Db 512 IleMetLysLeuProHisLysPheAspThrLeuValGlyGluArgIleAlaGlnLeuSer 531
OY 1831 GGAGAGAGAAAGCAAGAGTAGCAATTGCAGAGCCATTTTGAGAGACCCCGCAGTCATA 1890
Db 532 GlyGlyGlnLysGlnArgIleAlaIleAlaIleAlaValArgAsnProLysIleLeu 551
OY 1891 CTCTATGATGAAGCTACTTCATCGTTAGATTGATTAAGAGACTATTCTTGTCGCC 1950
Db 552 LeuLeuAspGluAlaThrSerAlaLeuAspThrGlnSerGluAlaValValGlnValAla 571
OY 1951 ATGAAGATGTGTCAACACAGACTTCTATTTCATTCGACACAGATTGCAACAGTG 2010
Db 572 LeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrVal 591
OY 2011 GTTGATGCAGATGAATCATTTGTTGGATCAGGTAGGTAGCCGAACTGTATCCAC 2070
Db 592 ArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGlnLysGlyAsnHis 611
OY 2071 CATGGTTTCCTGCTAACCTCATAGTATCTATTTCAGAAATGTGGCATACAGAGACG 2130
Db 612 AspGlnLeuMet---LysGlnLysSerGlyIleTyPheLysLeuValThrMetGlnThrAla 630
OY 2131 -----CGTGTGCAGAAC---CATGATTAACCCCAATGGAGACCAAGAAAGAA 2175
Db 631 GlyAsnGluValGlnLeuGlnAsnAlaAlaAspGluSerLysSerGlnIleAspAlaLeu 650
OY 2176 AATATATCCAAAGAGAGAAAGAAAGAACTA-----CAAGAGAAATTTGTCAATAGT 2229
Db 651 GluMetSerSerAsnAspSerArgSerSerLeuIleArgLysArgSerThrArgArgSer 670
OY 2230 GTGAAAGGC 2238
Db 671 ValArgGly 673
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2003, 03:23:25 ; Search time 65 Seconds  
(without alignments)  
9614.545 Million cell updates/sec

Title: AF133659  
Perfect score: 4254  
Sequence: 1 ATGGCGCTGCTGCGATGCA.....TATTAATAAATCATACATT 2345

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=framer\_n2p.model -BEV=xlh  
-G/cg2\_1/USPFO.spool/AF133659/runar.10022003.155440.29503/app\_query.fasta.1.2503  
-DB=A.Geneseq.101002 -OPMT=factan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=AF133659 @CGN.1.1.83 @runat.10022003.155440.29503 -NCPU=6 -ICPU=3  
-NO XLPXY -NO MAP -TAG=QUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: A.Geneseq.101002.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | % Match | Query Length | DB ID | Description |
|------------|--------|---------|--------------|-------|-------------|
| 1          | 3809   | 89.5    | 747          | 20    | AAW89585    |
| 2          | 3809   | 89.5    | 747          | 21    | AAAB1355    |
| 3          | 3736   | 87.8    | 752          | 20    | AAV14068    |
| 4          | 3656.5 | 86.0    | 722          | 22    | AAAB92941   |
| 5          | 1852   | 43.5    | 606          | 22    | ABBS7811    |
| 6          | 1852   | 43.5    | 606          | 22    | ABBS7812    |
| 7          | 1239.5 | 29.1    | 842          | 22    | AAU09965    |
| 8          | 1239.5 | 29.1    | 842          | 22    | AAE16764    |
| 9          | 1231   | 28.9    | 843          | 22    | AAU00010    |
| 10         | 1218   | 28.6    | 843          | 22    | AAU00011    |
| 11         | 1201.5 | 28.2    | 843          | 20    | AAV08961    |
| 12         | 1183   | 27.8    | 574          | 21    | AAV54454    |
| 13         | 1183   | 27.8    | 866          | 22    | ABB71210    |
| 14         | 1174   | 27.6    | 571          | 22    | AAAG3913    |
| 15         | 1149.5 | 27.0    | 591          | 22    | AAAG1099    |
| 16         | 964.5  | 22.7    | 342          | 21    | AAAG17365   |
| 17         | 920.5  | 21.6    | 327          | 21    | AAAG17366   |
| 18         | 889.5  | 20.9    | 319          | 21    | AAAG17367   |
| 19         | 739    | 17.4    | 147          | 22    | AAAG75566   |
| 20         | 736.5  | 17.3    | 640          | 23    | AAO22153    |
| 21         | 699.5  | 16.4    | 578          | 22    | AAU36908    |
| 22         | 695.5  | 16.3    | 582          | 23    | ABP39885    |
| 23         | 685    | 16.1    | 621          | 22    | AAAG92266   |
| 24         | 685    | 16.1    | 621          | 22    | AAAB7680    |
| 25         | 685    | 16.1    | 621          | 23    | AAE22051    |
| 26         | 682    | 16.0    | 548          | 22    | AAAG47516   |
| 27         | 668    | 15.7    | 577          | 23    | ABBA47542   |
| 28         | 654    | 15.4    | 582          | 22    | AAU38413    |
| 29         | 651    | 15.3    | 582          | 22    | AAU34522    |
| 30         | 647.5  | 15.2    | 543          | 22    | ABBA47270   |
| 31         | 643    | 15.1    | 593          | 23    | ABBA47177   |
| 32         | 641.5  | 15.1    | 707          | 18    | AAAB22153   |
| 33         | 641.5  | 15.1    | 707          | 21    | AAV51407    |
| 34         | 637.5  | 15.0    | 580          | 23    | ABPA40766   |
| 35         | 635    | 14.9    | 579          | 22    | AAAB96370   |
| 36         | 632    | 14.9    | 587          | 22    | AAU35367    |
| 37         | 632    | 14.9    | 587          | 22    | AAAB85504   |
| 38         | 632    | 14.9    | 587          | 23    | AAU91435    |
| 39         | 629.5  | 14.8    | 710          | 20    | AAV27214    |
| 40         | 629    | 14.8    | 710          | 20    | AAV27215    |
| 41         | 625    | 14.7    | 605          | 23    | ABBA47710   |
| 42         | 622    | 14.6    | 1279         | 22    | ABBB60234   |
| 43         | 621.5  | 14.6    | 583          | 22    | AAAB96741   |
| 44         | 621    | 14.6    | 1286         | 21    | AAAG39101   |
| 45         | 619    | 14.6    | 595          | 23    | ABP25492    |

## ALIGNMENTS

RESULT 1  
ID AAW89585 standard; Protein; 747 AA.  
AAW89585;  
22-MAR-1999 (first entry)  
Human ATP-binding cassette transport protein.  
Human, ATP-binding cassette transport protein; ABCCxH; cancer;  
neurological disorder; Alzheimer's disease; dementia; depression;  
Down's syndrome; epilepsy.  
Homo sapiens.  
US5858719-A.  
12-JAN-1999.

XX 17-JUL-1997; 97US-0895522.  
XX  
XX 17-JUL-1997; 97US-0895522.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Corley NC, Hillman JL, Shah P;  
XX  
DR WPI: 1999-119878/10.  
DR N-PSDB: AAV82520.  
XX  
PT Isolated ATP-binding cassette transport protein - used to develop  
PT products for treating cancers or neuronal disorders, e.g.  
PT Alzheimer's disease, dementia, depression, Down's syndrome or  
PT epilepsy  
XX  
PS Claim 1; Fig 1; 39pp; English.  
XX  
XX The present sequence represents human ATP-binding cassette transport  
XX protein (ABCTXH). The ABCTXH proteins are involved in disease-related  
XX transport processes, particularly in cancers and neuronal disorders.  
XX CC Antagonists of ABCTXH can be used to prevent or treat cancers or  
XX CC neuronal disorders e.g. Alzheimer's disease, amnesia, amyotrophic  
XX CC lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms,  
XX CC dementia, depression, Down's syndrome, tardive dyskinesia, dystonias,  
XX CC epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis,  
XX CC Parkinson's disease, paranoid psychoses, schizophrenia or Tourette's  
XX disorder.  
XX  
SQ Sequence 747 AA;  
XX  
Alignment Scores:  
Pred. No.: 0 Length: 747  
Score: 3809.00 Matches: 746  
Percent Similarity: 99.87% Conservative: 0  
Best Local Similarity: 99.87% Mismatches: 1  
Query Match: 89.54% Indels: 0  
DB: Gaps: 0  
AF133659 (1-2345) x AAW89585 (1-747)  
QY 16 ATGCATTTTGGCCGCTGGGCGCGCGGCGGCTTTCGAAAAGCGCGGCTCCGCG 75  
Db 1 MethisertPrArgTPrAlaAlaAlaAlaAlaPheGluLysArgRgiSeraLa 20  
QY 76 ATTGTGATCCGCGCTTGTAGTCTCTGTAGCGGCTCAGGTCGCGAGTGAAGCCACATCA 135  
Db 21 IleLeuIleArgProLeuValSerValSerGlySerGlyProGlnTPrArgProHisGln 40  
QY 136 CTCGGCGCTTGGGAAACCGCTCGAGCTTACAGATTCAGAGTCAATTAAGATATCA 195  
Db 41 LeuGlyAlaLeuGlyThrAlaArgAlaArgLlleProGlnSerLeuLysSerIleThr 60  
QY 196 TGGCAGAGATTTGGAAAAGCAATTCAGACAGATTCCTAGATGCTGCAAGGCTCCAG 255  
Db 61 TrpGlnArgLeuGlyLysGlyAsnSerGlyGlnPheLeuAspAlaAlaLysAlaLeuGln 80  
QY 256 GTATGGCCATGATGAAAGAGACATGTTGGCATGCTCATGACAGAGGAGACTCCAC 315  
Db 81 ValTrpProLeuIleGluLysArgThrCySerPheHisGlyHisAlaGlyLysGlyLeuHis 100  
QY 316 ACAGACCCAAAGAGAGGTAAAGATGTGATCTCGGAATATCATTAAGCAATGCTT 375  
Db 101 ThrAspProLysGluGlyLeuLysAspValAspPheArgLysIleLysAlaLysLeu 120  
QY 376 TCTTATGTGTGGCCCAAGACAGGCGAGATCTACAGCTAGAGTGGCATTTGCGTGGGA 435  
Db 121 SerTrpValTrpProLysAspArgProAspLeuArgAlaArgValAlaIleSerLeuGly 140  
QY 436 TTTTGGGTGGTGCAGAAAGCCATGAATTTGTGGTCCCTTCATGTTAAATATAGCTGA 495  
Db 141 PheLeuGlyGlyAlaLysAlaMetAsnIleValValProPheMetPheLysTyrAlaVal 160

QY 496 GACAGCCTCAACGAGATGTCGGGAAAATGCTGAACCTGATGATGACCAAAATACAGTT 555  
Db 161 AspSerLeuAsnGlnMetSerGlyAsnMetLeuAsnLeuSerAspAlaProAsnThrVal 180  
QY 556 GCACCACTGGCAACAGCATGTTGATGGCTATAGTGATCAAGAGCTGGAGCTGCTTTT 615  
Db 181 AlaThrMetAlaThrAlaValIleGlyTyrGlyValSerArgAlaGlyAlaAlaPhe 200  
QY 616 TTTTACGAGCTTGAATGCAATGAGATTTGGCAAGTGGCCGAAATTCATCCGAGATA 675  
Db 201 PheAsnGluValArgAsnAlaValPheGlyLysValAlaGlnMetSerIleArgGlyIle 220  
QY 676 GCCAAAATGCTCTTTCATCTTCATCTTCACAACTGATCTGGGTTTTCACCTGACAGACAG 735  
Db 221 AlaLysAsnValPheLeuHisLeuHisAsnLeuAspLeuGlyPheHisLeuSerArgGln 240  
QY 736 ACGGAGACTTATCTAAAGCTATTCACAGAGCAAGAGGATTCAGTTTGTCTGAGT 795  
Db 241 ThrGlyAlaLeuSerLysAlaIleAspArgGlyThrArgGlyIleSerPheValLeuSer 260  
QY 796 GCTTTGATTTATATCTTCTCCATCAGTTCGATGATGATGATGATGATGATGATGATG 855  
Db 261 AlaLeuValPheAsnLeuLeuProIleMetPheGluValMetLeuValSerGlyValLeu 280  
QY 856 TATTACAATGCGCGTCCAGTTGGCTTTGGTAAACCTTGGAAACACTTGGTACATACACA 915  
Db 281 TyrTyrLysSerGlyLysGlnPheAlaLeuValThrLeuGlyThrLeuGlyThrTyrThr 300  
QY 916 GCATTCACAGTTGAGTGCACAGGTCGAGCACTGATTTGAATGAAATGAACAAAGCA 975  
Db 301 AlaPheThrValAlaValThrArgTPrArgThrArgPheArgIleGluMetAsnLysAla 320  
QY 976 GATATATGATCAGGATATGCTGCTATAGACTCATGCTGAATATGAACTGGAAGTAT 1035  
Db 321 AspAsnAspAlaGlyAsnAlaAlaIleAspSerLeuLeuAsnLysGlnThrValLysTyr 340  
QY 1036 TTTAATATGAAAGATATGAGACAGACAGATATGATGATTTTGAAGAGTATGAGACT 1095  
Db 341 PheAsnAsnGlnArgTyrGluAlaGlnArgTyrAspGlyPheLeuLysThrTyrGluThr 360  
QY 1096 GCTTCATGAAAGTACCTCTACTCTGCTATGCTGAACTTGGTCAAAAGCTATTTTC 1155  
Db 361 AlaSerLeuLysSerThrSerThrLeuAlaMetLeuAsnPheGlyGlnSerAlaIlePhe 380  
QY 1156 AGTGTGGTTTAACTGCTAAATAGTGCTCCGACAGAGAAATTTGGAGGTTACCTT 1215  
Db 381 SerValGlyLeuThrAlaIleMetValLeuAlaSerGlnGlyIleValAlaGlyThrLeu 400  
QY 1216 ACTGTTGAGATCTAGTAATGAGATGAGACTGCTTTTTCAGCTTCATTAACCTGAAAC 1275  
Db 401 ThrValGlyAspLeuValMetValAsnGlyLeuLeuPheGlnLeuSerLeuProLeuAsn 420  
QY 1276 TTTCTGGAACTGTATATAGAGACTTACAGACACTCATATGAAATGAAACACTTGT 1335  
Db 421 PheLeuGlyThrValTyrArgGlyThrArgGlnAlaLeuIleAspMetAsnThrLeuPhe 440  
QY 1336 ACTCTACTCAAGGTAGACCCCAATTAAGACAAAGTGAATGGCACTCCCTTCAGATC 1395  
Db 441 ThrLeuLeuLysValAspThrGlnIleLysAspLysValMetAlaSerProLeuGlnIle 460  
QY 1396 ACACACAGACAGCTACCGGCTTGTGATGATGATGATGATGATGATGATGATGATG 1455  
Db 461 ThrProGlnThrAlaThrValAlaPheAspAsnValHisPheGluTyrIleGluGlyGln 480  
QY 1456 AAGTCTTATAGTGAATATCTTGAAGTCCCTGACAGAAAGAAAGTGGCCATTTGAGA 1515  
Db 481 LysValLeuSerGlyLysSerPheGluValProAlaGlyLysLysValAlaIleValAla 500  
QY 1516 GGTAGTGGTCAAGGAAAGCAATAGTGAAGCTATTTTCGCTTATGAGCCTCAA 1575  
Db 501 GlySerGlySerGlyLysSerThrIleValArgLeuLeuPheArgPheTyrGluProGln 520

OY 1576 AAGGTAGCATTTATCTTGGTCAAAATATACAAAGTGAAGCTGGAAACCTTCGG 1635  
DB 521 LysGlySerIleTyrLeuAlaGlyGlnAsnIleGlnAspValSerLeuGlnSerLeuArg 540  
OY 1636 AAGGACGTGGAGTGTACTGAGGATGCTGCTTCCATTAATCTTTATTAACAAC 1695  
DB 541 ArgAlaValGlyValValProGlnAspAlaValLeuPheHisAsnThrIleTyrTyrAsn 560  
OY 1696 CTCTTATATGAAACATCAGTCTTCACCTGAGAAAGTATGACAGTGGCAAAATAGCT 1755  
DB 561 LeuLeuTyrGlyAsnIleSerIleSerProGlnGlnValTyrAlaValAlaLysLeuAla 580  
OY 1756 GCACTTCATGATGCAATTTCTTGCAATGCCACATGATATGACACCCAGTAGGGGAACGA 1815  
DB 581 GlyLeuHisAspAlaIleLeuArgMetProHisGlyTyrAspThrGlnValGlyIuArg 600  
OY 1816 GGAAGTCAAGCTTTCAGAGAGAAAGAAAGAGATGACAAATGGCAAGCCATTTTGAAG 1875  
DB 601 GlyLeuLysLeuSerGlyGlyGlnArgValAlaIleAlaArgAlaIleLeuLys 620  
OY 1876 GACCCCGCAGTCATCTCTATGATGAGTACTTGCATCGTTAGATTGCAATTAAGAG 1935  
DB 621 AspProProValIleLeuTyrTrpSerIuAlaThrSerSerLeuAspSerIleThrGln 640  
OY 1936 ACTATTTCTGGTCCATGGAAGATGTGTCAACAACAGAACTTCTATTTCATTCGACAC 1995  
DB 641 ThrIleLeuGlyAlaMetLysAspValValLysHisArgThrSerIlePheIleAlaHis 660  
OY 1996 AATTTGTCAACGTGTGATGATGCAATGAAATCATTTGTTTGGATCAGGGTAAAGTAGCC 2055  
DB 661 ArgLeuSerThrValValAlaAspAlaAspGlnIleIleValLeuAspGlnIleLysValAla 680  
OY 2056 GAACGTGGTACCCACATGATGTTGCTGTGTAACCTCATGATATCTTATTCAGAAATGTGG 2115  
DB 681 GluArgGlyThrHisHisGlyLeuLeuAlaAsnProHisSerIleTyrSerGlnMetTrp 700  
OY 2116 CATACACAGAGACCGGTGTGCAAGAACCATGATTAACCCCAATGGGAACAAAGAAAGA 2175  
DB 701 HistTrgInSerSerArgValGlnAsnHisAspAsnProLysTrpGlnAlaLysGln 720  
OY 2176 AATATATCCAAAGAGAGAAAGAAAGAAACTACAAAGAAATTTGTCAATAGTGTGAAA 2235  
DB 721 AsnIleSerLysGlnGlnGlnIuArgLysLysLeuGlnGlnIuIleValAsnSerValLys 740  
OY 2236 GGCTGTGAAAACCTGTTGCTGC 2256  
DB 741 GlyCysGlyAsnCysSerCys 747  
RESULT 2  
AAB13355  
ID AAB13355 standard; Protein; 747 AA.  
XX  
AC AAB13355;  
XX  
DT 13-NOV-2000 (first entry)  
XX  
DE Human ATP-binding cassette transport protein.  
XX  
KW Human; ATP-binding cassette transport protein; ABCtxH; cytosolic;  
KW neurotrophic; neuroprotective; cerebroprotective; antidepressant;  
KW anticonvulsant; antiparkinsonian; neuroleptic; cancer; epilepsy;  
KW neurodegenerative disorder; akathisia; amyotrophic lateral sclerosis;  
KW bipolar disorder; catatonia; dementia; depression; Down's syndrome;  
KW tardive dyskinesia; dystonia; multiple sclerosis; neurofibromatosis;  
KW schizophrenia; Tourette's disease.  
OS Homo sapiens.  
XX  
PN US608042-A.  
XX  
PD 27-JUN-2000.  
XX  
PF 18-NOV-1998; 98US-0195391.

XX 17-JUL-1997; 97US-0895522.  
PR  
XX  
XX (INCY-) INCYTE PHARM INC.  
PA  
XX  
XX Shah P, Corley NC, Hillman JL;  
PI  
XX WPI; 2000-451228/39.  
DR  
XX N-PSDB; AAA62411.  
PT  
XX  
XX New human ATP-binding cassette transport protein useful for diagnosing,  
PT preventing and treating cancers, e.g. leukemia, sarcoma, and neuronal  
PT disorders, e.g. schizophrenia, dementia  
XX  
XX  
PS Claim 1; Fig 1; 31pp; English.  
XX  
XX The present sequence is a human ATP-binding cassette transport protein  
CC (ABCTxH). The nucleotide sequence encoding ABCTxH was isolated from a  
CC human ovarian tissue cDNA library. Clones from the library were sequenced  
CC and used as query sequences against homology databases. ABCTxH encoding  
CC polynucleotides were extended using PCR. ABCTxH polynucleotides and  
CC polypeptides may be used for the diagnosis, prevention of ABCTxH, e.g.  
CC conditions or disorders associated with the expression of ABCTxH, e.g.  
CC adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, akathisia, Alzheimer's disease, amnesia, amyotrophic  
CC lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms,  
CC dementia, depression, Down's syndrome, tardive dyskinesia, dystonias,  
CC epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis,  
CC Parkinson's disease, paranoid psychoses, schizophrenia, and Tourette's  
CC disease.  
XX  
SQ Sequence 747 AA;  
Alignment Scores:  
Pred. No.: 0 Length: 747  
Score: 3809.00 Matches: 746  
Percent Similarity: 99.87% Conservative: 0  
Best Local Similarity: 99.87% Mismatches: 1  
Query Match: 89.54% Indels: 0  
DB: 21 Gaps: 0  
AF133659 (1-2345) x AAB13355 (1-747)  
OY 16 ATGCATCTTGGCCGTGGCGCGCGCGCTGCTTCAAAAAGCGCGGACATCCGCG 75  
DB 1 MethSerTrpArgTrpAlaAlaAlaAlaAlaAlaPheGlnLysArgArgHisSerAla 20  
OY 76 ATTCTGATCCGCGCTTATAGTCTCTGTAGCGGCTCAGGTCGCGAGTGGAGGCCATGAA 135  
DB 21 IleLeuIleArgProLeuValSerValSerGlySerGlyProGlnTrpArgProHisGln 40  
OY 136 CTGGCGCGCTTGGAGACCGCTCGAGCTTACCAAGTTCAGAGTCAATTAAGATACCA 195  
DB 41 LeuGlyAlaLeuGlyTrpAlaArgAlaTyrGlnIleProGlnSerLeuLysSerIleThr 60  
OY 196 TGGCAGAGATTGGGAAAAGCAATTCAGACAGTTCTTAGATGGCGAAAGCTCTCAG 255  
DB 61 TrpGlnArgLeuGlyLysGlyAsnSerGlyGlnPheLeuAspAlaAlaLysAlaLeuGln 80  
OY 256 GTATGGCCATGATAGAAAAGAGACATGTTGGCATGGTCATGCAAGAGAGACTCCAC 315  
DB 81 ValTrpProLeuIleGlnLysArgTrpCysTrpHisGlyHisAlaGlyGlyLeuHis 100  
OY 316 ACAAGCCAAAAGAAAGGTTTAAAGATGTGATCTCGGAAAATCATTAAGCAATGCTT 375  
DB 101 ThrAspProLysGlnGlyLeuLysAspValAspThrArgLysIleIleLysAlaLysLeu 120  
OY 376 TCTATGATGAGCCCAAGACAGCCAGATCTAGAGTAGAGTGGCATTTGCTGGGA 435  
DB 121 SerTyrValTrpProLysAspArgProAspLeuArgAlaArgValAlaIleSerLeuGly 140  
OY 436 TTTTGGGTGTGCAAGGCCATGATATTTGTGTTCCCTTCATGTTAAATATGCTGTA 495

Db 141 PheLeuGIyGIyAlaIySaIaMeIaEnIleValIaIProPheMeCPheUySTyRaIaVal 160  
 QY 496 GACAGCTCAACCAAGATGTCGGGAAACATGCTGACCTGATGACCAAAATTAACGTT 555  
 Db 161 AspSerLeuAsnGImetSerGIyAsnMeIeAsnLeuSerSpsIaIaProAsnTrVal 180  
 QY 556 GCAACCATGGCAACAGAGCTTCTGATTTGGCTATGCTATCAAGAGCTGGAGCTGCTTTT 615  
 Db 181 AlaThrMeAlaThrAlaValIleuIleGIyTyGIyValSerArgIaGIyAlaIaPhe 200  
 QY 616 TTTTAACGAAGTTGAAATGACGATTTTGGCAAGTACCCAGAAATTCATCCGAAGAT 675  
 Db 201 PheAsnGIuValaArgAsnIaValaPheGIyLySaIaIaGIaAsnSerIleAsnArgIle 220  
 QY 676 GCAAAATATGCTTCTCCATCTTCAACCTGATCTGGGTTTCACTGAGCAGACAG 735  
 Db 221 AlaIySaIaIaPheLeuHISpLeuHISpLeuAspLeuGIyPheHISpLeuSerArgGln 240  
 QY 736 ACGGAGCTTTATCTAAGCTATTGACAGAGAAAGGGGTATCAGTTTGTCTGAGT 795  
 Db 241 ThrGIyAlaLeuSerIySaIaIleAsnArgGIyThrArgGIyIleSerPheValIleuSer 260  
 QY 796 GCTTTGGATTTAATCTTCTCCATCATGTTTGAAGTGAAGTCTTGTCAGTGGTGT 855  
 Db 261 AlaIeUValaPheAsnLeuLeuProIleMeCPheGIuValaMeIeUValaISerGIyValIleu 280  
 QY 856 TTTTAACAAATGGGTCGCCAGTTTGGTTGGTAAACCTTGGACACTGGTATACAC 915  
 Db 281 TyTyTyLyCySeGIyAlaGlnPheAlaIeUValaThrLeuGIyThrLeuGIyTrpH 300  
 QY 916 GCATTCAAGTTGACAGTCAACCGGTGAGAACTAGATTTAGAAATAGAAATGAACAAAGCA 975  
 Db 301 AlaPheThrValaIaValaIaThrArgTrpArgThrArgPheArgIleGIuMeIaAsnLySaIa 320  
 QY 976 GATATGATGACAGTATGCTGCTATGACTCACTGCTGAATTAATGAAGTGAAGTAT 1035  
 Db 321 AspAsnAspAlaGIyAsnAlaIaIleAspSerLeuAsnTrpGIuThrValaIySTy 340  
 QY 1036 TTTTAATGAAGATATGAGACAGAGATATGATGATTTTGAAGCGTATGAGACT 1095  
 Db 341 PheAsnAsnGIuArgTrpGIuAlaGlnArgTyAspGIyPheLeuYsnTrpTyGIuTrpH 360  
 QY 1096 GCTTCATGAAAAGTACCTTACTGCTGCTATGCTGAACTTGGTCAAAAGTCTATTTTC 1155  
 Db 361 AlaSerLeuIySerThrSerThrLeuAlaMeIeUAsnPheGIyGlnSerAlaIlePhe 380  
 QY 1156 AGTGTGGTTTAAACAGCTATATGCTGCTGCCAGTCAAGGAATTTGGCAGGTACCTT 1215  
 Db 381 SerValGIyLeuThrAlaIleMeIeValIleuAlaSerGlnGIyIleValaIaGIyThrLeu 400  
 QY 1216 ACTGTTGGAGTCTAGTAATGGAATGACCTGTTTTCAGTTTCATTACCCCTGAAC 1275  
 Db 401 ThrValGIyAspLeuValMeIeValaAsnGIyLeuLeuPheGlnIleuSerLeuProLeuAsn 420  
 QY 1276 TTTCTGGAACTGATATAGAGAGACTAGACAAGCACTATAGATATGAACCTTGT 1335  
 Db 421 PheLeuGIyThrValaIyTrpArgGIuThrArgGlnAlaLeuIleAspPheAsnThrLeuPhe 440  
 QY 1336 ACTCTACTCAAGGTAGACACCAAAATTAAGACAAAGTGCATCTCCCTTCAGATC 1395  
 Db 441 ThrLeuLeuIyValaIaPheTrpGlnIleIyAspLySaIaIaMeIaIaSerProLeuGlnIle 460  
 QY 1396 AACAACACAGACAGTACCGTGGCTTGAATGTGATTTTGAATTAATCAATTAAGGGCCAG 1455  
 Db 461 ThrProGlnThrAlaThrValaIaPheAspAsnValaHISpHeGIuTyTrpIleGIuGIyGln 480  
 QY 1456 AAAGTCTTAGTGAATATCTTTGAAGTCCCTGCAGGAAAGAAAGTGGCACTTGTAGA 1515  
 Db 481 LysValaIleuSerGIyIleSerPheGIuValaProIaGIyLySaIaIaIleValaGIy 500  
 QY 1516 GGTAGTGGGTCAAGGAAAGACAAATAGTGAAGCTATTAATTTGCTTATGAAGCTCA 1575  
 Db 501 GlySerGIySerGIyLysSerThrIleValaArgLeuLeuPheArgPheTyArgIuProGln 520

QY 1576 AAGGTAGCATTTATCTTCTGCTGTCAAAATATACAGATGTGAGCTGGAAAGCTTCGG 1635  
 Db 521 LysGIySerIleIeTyLeuAlaGIyGlnAsnIleGlnAspValSerLeuGlnSerLeuArg 540  
 QY 1636 AGGGCAGTGGAGTGTACCTCAGAGATGCTGTCTTCTTCATTAATCAATTAATTAAC 1695  
 Db 541 ArgAlaValaGIyValaIaProGlnAspAlaValaIleuPheHISpAsnThrIleTyTrpAsn 560  
 QY 1696 CTCTTAATNGAACAATCAAGTCTTCACTGAGGAAGGTATGACAGCCCAATGAGTAACTAGCT 1755  
 Db 561 LeuLeuTyGIyAsnIleSerAlaSerProGIuGIuValaTyAlaValaIaLySaIa 580  
 QY 1756 GCACTTCATGATGCAATTTCTTGAATGCGCACATGATATGACACCCCAATGAGGAAAGCA 1815  
 Db 581 GlyLeuHISpAlaIleIleuArgMeCPheHISpGIyTyTrpAspThrGlnValaGIyIuArg 600  
 QY 1816 GCACTCAAGCTTTCAGAGAGAGAAAGCAAGAGTATGCAATTTGCAAGCCATTTTGAAG 1875  
 Db 601 GlyLeuLySaIeSerGIyGIyGIuLySaIaArgValaIaIleAlaArgAlaIleLeuLyS 620  
 QY 1876 GACCCCGCATCTCATATGATGAGCTACTGATGCTTATGATTCATTAAGTCAAGAG 1935  
 Db 621 AspProIValaIleLeuTyAspGIuAlaTrpSerSerLeuAspSerIleThrGIuGIu 640  
 QY 1936 ACTATTTGTGTCATGAGAGATGTGTCAAACACAGAACTTCTAATTTTCATTGCACAC 1995  
 Db 641 ThrIleLeuGIyAlaMeIeLySaIaValaIySbHISpArgThrSerIlePheIleAlaHIS 660  
 QY 1996 AGATTGTCAACAGTGTGATGACAGATGAATCATTTGCTGATCAGGGTAAAGTACCC 2055  
 Db 661 ArgLeuSerThrValaIaAspAlaAspGlnIleIleValaLeuAspGlnGIyLySaIaIa 680  
 QY 2056 GAACTGTGAACCAACAGCTTTCCTGCTAACCTCATAGATCATTAATCAGAAATGTGG 2115  
 Db 681 GluArgGIyThrHISpIleGIyLeuLeuAlaAsnProHISpSerIeTySerGIuMeIeTrp 700  
 QY 2116 CATPACAGAGCAGCGGTGTCAGAACCATGATTAACCCCAATGGGAAAGCAAGAAAGCA 2175  
 Db 701 HISpGlnSerSerArgValaGlnAsnHISpAsnProLyTrpGIuAlaLySaIySgIu 720  
 QY 2176 AATATATCCAAAGAGAGAGAAAGAAAGAACTPACAAAGAAATTTGCTAATGTGTGAAA 2235  
 Db 721 AsnIleSerLyGIuGIuGIuArgLySbLySbLeuGlnGIuGIuIleValaAsnSerValaLyS 740  
 QY 2236 GGCTGTGAAACTGTTCGTGC 2256  
 Db 741 GlyCySeGIyAsnCySerCyS 747  
 RESULT 3  
 AAY14068  
 ID AAY14068 standard; Protein; 752 AA.  
 AC AAY14068;  
 XX  
 DT 16-JUL-1999 (first entry)  
 DT XX  
 DE Human ABC-Transporter-7 (HABCT), protein sequence.  
 XX  
 KW HABCT; human; ABC-transporter-7; diagnosis; cancer; autoimmune disease;  
 KW Addison's disease; insulin-dependent diabetes mellitus; therapy;  
 KW microsomal disorder.  
 OS Homo sapiens.  
 OS XX  
 PN WO9921885-A1.  
 PD 06-MAY-1999.  
 PD XX  
 PF 29-OCT-1997; 97WO-CN00120.  
 PF XX  
 PR 29-OCT-1997; 97WO-CN00120.  
 PR XX



PA (UYSH-) UNIV SHANGHAI SECOND MEDICAL.

PI Zhang Q;

DR WPI; 1999-312945/26.

DR N-PSDB;

PT HABC7 a protein useful in the treatment of cancer, diabetes, PT microsomal disorders and Addison's disease

PS Claim 11; Page 27-29; 31pp; English.

This sentence is the human ABC-transporter-7 (HABCT) protein of the invention. HABCT is an ABC transporter protein. The DNA, vectors containing it and recombinant host cells are useful for recombinant production of HABCT. The DNA, HABCT and antibodies against HABCT are useful as research reagents, for screening assays and in diagnostic assays. Antagonists and agonists of HABCT can be used to inhibit or enhance, respectively, the activity of HABCT or expression of the HABCT coding sequence. Anti-HABCT antibodies and HABCT or its fragments can be used in vaccines. In particular, the proteins, antibodies, agonists and antagonists can be used for treating, e.g., cancer, autoimmune disease, Addison's disease, microsomal disorders and insulin-dependent diabetes mellitus, related to both an excess and insufficient amounts of HABCT.

**SQ** Sequence 752 AA;

### Alignment Scores:

|                        |         |               |
|------------------------|---------|---------------|
| Pred. No.:             | 0       | 752           |
| Score:                 | 3736.00 | Matches:      |
| Percent Similarity:    | 98.54%  | Conservative: |
| Best Local Similarity: | 98.01%  | Mismatches:   |
| Query Match:           | 87.82%  | Indels:       |
| DB:                    | 20      | Gaps:         |
|                        |         | 2             |
|                        |         | 2             |

AF133659 (1-2345) X AAY14068 (1-752)

[illegible]

|    |      |  |      |
|----|------|--|------|
| QY | 541  | GACGCAATKACAGTGAACAACAAAGGACAGAGCTGTGATTGGCATATGGTTATACAGA     | 600  |
| Db | 181  | AAlaProsmnThrValAlaThrMetAlaThrAlaValLeuIleIlyTrIyIalValSerArg | 200  |
| QY | 601  | GCTGACACTCCTTTTTTAACGAAGTTCGAATGCAGTATTTTGSCAAGGTACCCAGAA      | 660  |
| Db | 201  | AlaGlyLalaIaPhePheAsnGluValArgAsnAlaValPheGluYusValAlaGlnAsn   | 220  |
| QY | 661  | TCAATCCGAAGAAATAGCCAAAATAATGCTTTCTCCATCTTCAACACTGGATCTGGTTT    | 720  |
| Db | 221  | SerIleArgArgIleAlaYusAsnValPheLeuHIsLeuHIsAsnLeuAspLeuGlyPhe   | 240  |
| QY | 721  | CACCTGACGACAGACAGCGGAGCTTATATGATAGGCTATGACAGAGGAAACAAGGGTATC   | 780  |
| Db | 241  | HIsLeuSerArgGlnThrGlyAlaLeuSerLySAlaIeAspArgGlyThrArgGlyIle    | 260  |
| QY | 781  | AGTTTCTCTGAGTCTGTTGGATTTAACTCTTCCCATCATGTTTGAAGTAGT---         | 837  |
| Db | 261  | SerPheValLeuSerAlaLeuValPheAsnProLeuProAsnHIsValGluValMetLeu   | 280  |
| QY | 838  | CTTGTCAGTGTGTTTGTATTAACAATCGCGTCCAGTTTGCTTTGGTAAACCTTGA        | 897  |
| Db | 281  | LeuValSerGlyAlaLeuTryrTrpScysCysAlaGlnLeu--LeuGluAsnLeuGly     | 299  |
| QY | 898  | ACACTTGGTATATACACAGATTCACAGTTTGCAGTACACCGGTGAGAACTGATTTAA      | 957  |
| Db | 300  | ThrLeuGluYThrTrpThrAlaPheThrValAlaValaThrArgTrpArgThrArgPheArg | 319  |
| QY | 958  | ATGAAATGAAACAAGACAGTATATGATGACAGGAAATGCTCATAGACTCACTGGCAT      | 1011 |
| Db | 320  | LeuGluIleAspGlnAlaAspAsnAspAlaGlyAsnAlaAlaIleAspSerLeuLeuAsn   | 339  |
| QY | 1018 | TATCAAACCTGTGAAGATTTTATATATGAAGATATGAAGCACAGAGATATGATGATT      | 1077 |
| Db | 340  | TyrGlnThrValIleYsrYrPheAsnAsnGluArgTrpGluAlaGlnArgTrpAspGlyPhe | 359  |
| QY | 1078 | TTGAGACGATAGACACTGCTTCATTGAAAGTACTCTTACTCTGGCTATGCTGAAC        | 1133 |
| Db | 360  | LeuYsrThrTrpGlnThrAlaSerLeuYsrSerThrSerThrLeuAlaMetLeuAsnPro   | 379  |
| QY | 1138 | GGTCAAAAGTCTATTTTTCAGTGCCTGTTTAAACAGCATATATGGTGTGCCAGTACGGA    | 1197 |
| Db | 380  | GlyGlnSerAlaIlePheSerValGlyLeuThrAlaIleMetValLeuAlaSerGlnGly   | 399  |
| QY | 1198 | ATTGTGGCAGAGTACCCTTACTGTGTGAGATCTAGTATGTCATATGACTGCTTTTTCAG    | 1257 |
| Db | 400  | IleValAlaIleYrThrLeuThrValGlyAspLeuValMetValAsnGlyLeuLeuPheGln | 419  |
| QY | 1258 | CTTTCATTACCCCTGAACTTCTGGGAACGTATATAGAGACTTACAGAACACTCATAT      | 1317 |
| Db | 420  | LeuSerLeuProLeuAsnPheLeuGlyThrValYrArgGluThrArgGlnAlaLeuIle    | 439  |
| QY | 1318 | GATATGAACACTTGTTTACTCTACTCAAGGTATAGACACCCCAATTTAAGACAAAGTGATG  | 1377 |
| Db | 440  | AspMetAsnThrLeuPheThrLeuLeuYsValAspThrGlnIleYsAspYsValMet      | 459  |
| QY | 1378 | GCAATCCCCCTTCAATCACACACAGACGATCCGTGGCTTTGATATATGCAATTT         | 1437 |
| Db | 460  | AlaSerProLeuGlnIleThrProGlnThrAlaThrValAlaPheAsnValHIsPhe      | 479  |
| QY | 1438 | GAATACATTGAGGGCCAGAAAGTCTTATGGAATATCTTTGAAGTCCCTGCAGAGAA       | 1497 |
| Db | 480  | GluTrpIleGluGlyGlnIleYsValLeuSerGlyIleSerPheGluValProHlaGlyYs  | 499  |
| QY | 1498 | AAAGTGGCCATTGTAGAGGTAGTGGGTAGGGAAGAACACATATGTAGGCTATTATT       | 1557 |
| Db | 500  | LysValAlaIleValIleGlySerGlySerGlyYsSerThrIleValaArgLeuLeuPhe   | 519  |
| QY | 1558 | CGCTTCATAGAGCTCAAAAAGGTAGCATTTATCTTGCTGCGCAAAATTTACAAAGTGG     | 1617 |
| Db | 520  | ArgPheTrpGlnProGlnYsGlySerIleYrIleAlaIleGlyGlnAsnIleGlnAspVal  | 539  |
| QY | 1618 | AGCTGGAAAGCCTTGCAGAGGACATGGGAGATGGTAACTCAGAGATGCTGCTCTTTCAT    | 1677 |

```
Db 540 SerLeuGluSerLeuArgAlaValAlaValProGlnAspAlaValLeuPheHis 559
Oy 1678 AATACTATTATTATCAACCTCTTATATGGAACATCAGCTTCCCTCAGGAAGATGAT 1737
Db 560 AsnThrIleTyrTrpAsnLeuLeuTyrGlyAsnIleSerAlaSerProGlnValValTyr 579
Oy 1728 GCAGTGGCAAAATTAGTGCATCTTCATGATGCAATTTCTTGAATGCCATGATGATGAC 1797
Db 560 AlaValAlaValSerLeuAlaGlyLeuHisAspAlaIleLeuAlaGlyMetProHisGlyTyrAsp 599
Oy 1798 ACCCAAGTAGGGGAAAGAGAGACTCAAGCTTTGAGAGAGAGAAAGCAAGATGCAAT 1857
Db 600 ThrGlnValAlaGlyAlaGlyLeuLeuSerGlyGlyGlyValAlaIle 619
Oy 1858 GCAAGAGCATTTTGAAGAGACCCCGCATACCTCTATGATGAAGACTCTTCATCTGTA 1917
Db 620 AlaArgAlaIleLeuLysAspProProValIleLeuTyrAspGlyAlaThrSerSerLeu 639
Oy 1918 GATTGATTACTGAAGAGACTATTCTTGTCATGAAGAGATGGTCAAAACAGAACT 1977
Db 640 AspSerIleThrGlnGlnIleThrIleLeuGlyAlaMetLysAspValValLysHisArgThr 659
Oy 1978 TCTATTTTCATTGCACACAGATTGTCAACAGTGTGATGACAGATGAATCATTTGTTG 2037
Db 660 SerIlePheIleAlaHisArgLeuSerThrValAlaAspGlnIleIleValLeu 679
Oy 2038 GATCAGGGTAAAGTACCCCAAGCTGTGATCCCAACATGTTGCTTCTCCTCATACT 2097
Db 680 ArgGlnGlyLysValAlaGlyAlaGlyLysThrIleHisGlyLeuLeuAlaAsnProHisSer 699
Oy 2098 ATCTATTGCAATGTGCGATACACAGAGACCGGTGTCAGAACCATGATTAACCCCAA 2157
Db 700 IleTyrSerGlnMetIleThrHisThrGlnSerSerArgValGlnAsnHisAspAsnProLys 719
Oy 2158 TGGGAAGCAAAAGAAAATAATATTCAAAGAGAGAGAGAAAGAACTACAGAGAA 2217
Db 720 TrpGlnAlaLysLysGlnLysHisSerLysGlnGlnAlaArgLysLeuGlnGln 739
Oy 2218 ATTGTCAATAGTGTGAAGAGCTGTGAAACTGTTCCTGTC 2256
Db 740 IleValAsnSerValLysGlyCysGlyAsnCySerCys 752

RESULT 4
AAB92941
ID AAB92941 standard; Protein; 722 AA.
AC AAB92941;
XX 26-JUN-2001 (first entry)
DT
XX
DE Human protein sequence SEQ ID NO:11610.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
```

```
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11610; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to a
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 722 AA;
XX

Alignment Scores:
Pred. No.: 0 Length: 722
Score: 3656.50 Matches: 720
Percent Similarity: 99.72% Conservative: 0
Best Local Similarity: 99.72% Mismatches: 1
Query Match: 85.95% Indels: 1
DB: 22 Gaps: 1

AF133659 (1-2345) x AAB92941 (1-722)
Oy 1 ATGGGGCTGCTCGCATGATCTTTGGCGCTGGCGCGCGCTTTCGAAAG 60
Db 1 MetAlaLeuLeuAlaMetHisSerTyrArgTrpAlaAlaAlaAlaPheGluLys 20
Oy 61 CGCCGCACTCCGCGATTCGTGATCCGCGCTTAGTCTGTAGCGCGCTCAGGTCGAG 120
Db 21 ArgArgHisSerAlaIleLeuIleArgProLeuValSerValSerGlySerGlyProGln 40
Oy 121 TGGAGGCCATCAACTGGCGGCTTGGGAACCGCTGACGCTTAC--CAGATTCACAG 177
Db 41 TrpArgProHisGlnLeuGlyAlaLeuGlyThrAlaArgAlaTyrGlnGlnIleProGln 60
Oy 178 TCAATMAAAGATATCACATGCGAGATTTGGGAAGCAATTCAGACAGTCTTAGAT 237
Db 61 SerLeuLysSerIleThrTrpGlnArgLeuGlyLysGlyAsnSerGlyGlnPheLeuAsp 80
Oy 238 GCTGCAAAAGGCTCTCCAGGTAAGCCACTGATAGAAAAGAGACATGTTGGCATGTCAT 297
Db 81 AlaAlaLysValAlaLeuGlnValTrpProLeuIleGlyLysArgThrCysTrpHisGlyHis 100
Oy 296 GCAGAGAGAGAGCTCCACACAGACCCCAAGAGGCTTAAAGATGTTGATCTCGGAAA 357
Db 101 AlaGlyGlyGlyLeuHisThrAspProLysGlyGlyLeuLysAspValAlaAspThrArgLys 120
Oy 358 ATCATMAAAGCAATGCTTTCTTATGTGTGGCCCAAGACAGGCCGATCTACGAGCTGA 417
Db 121 IleIleLysAlaMetLeuSerTyrValTrpProLysAspArgProAspLeuArgAlaArg 140
```

Qy 418 GTTGCAATTTGGCTGGGATTTTGGGTGGTCAAAAGCCATGAATATGTGGTTCCTTC 477  
 Db 141 ValAlaIleSerLeuGlyPheLeuGlyGlyValAlaYsaIleMetAsnIleValValProPhe 160  
 Qy 478 ATGTTAAATATGCTGTAGACAGCCTCAACCAAGATGTGGGAAACATGCTGAACCTGAGT 537  
 Db 161 MetPheYsrYrAlaValAspSerLeuAsnGlnMetSerGlyAsnMetLeuAsnLeuSer 180  
 Qy 538 GATGCACCAAAATAGACGTTGCAACCAATGGCAACAGCAGTTCGTATGGCTATGGTATCA 597  
 Db 181 AspAlaProAsnThrValAlaThrMetAlaThrAlaValLeuIleGlyTyrGlyValSer 200  
 Qy 598 AGAGCTGAGCTGCTTTTATTAAGCAAGTTGAAATGCAATTTGGCAAGCTAGCCAG 657  
 Db 201 ArgAlaGlyAlaAlaPhePheAsnGlyValAlaArgAsnAlaValPheGlyValAlaGln 220  
 Qy 658 AATTCAATCCGAGAAATAGCCAAAAATGCTCTTCCATCTTCCATCAACCTGAGTCTGGCT 717  
 Db 221 AsnSerIleArgArgIleAlaAlaYsaAsnValPheLeuHISLeuHISAsnLeuAspLeuGly 240  
 Qy 718 TTTGACCTGACAGACAGACCGGAGCTTTATCTAAGCTATGACAGAGAAAGAGGT 777  
 Db 241 PheHisLeuSerArgGlnThrGlyAlaLeuSerYsaAlaIleAspArgGlyThrArgGly 260  
 Qy 778 ATCAGTTTGTCTGAGTCTTTGGATTTAATCTTCTCCATCATGTTGAAGTGAATG 837  
 Db 261 IleSerPheValIleuSerAlaLeuValPheAsnLeuLeuProIleMetPheGlyValMet 280  
 Qy 838 CTGTGCACTGGTGTCTTTGATTATTAACAATGGGTGCCAGTTTGGTGGTGAACCTTGGGA 897  
 Db 281 LeuValSerGlyValLeuYrYrLysGlySglYalagInPheAlaLeuValThrLeuGly 300  
 Qy 898 ACACCTGGTACATACAGACATTCACAGTTGCAGTCAACCGGTGAGAGATTAATTAAGA 957  
 Db 301 ThrLeuGlyThrYrThrAlaPheThrValAlaValAlaThrArgThrArgPheArg 320  
 Qy 958 ATAGAATGAACAAAGCAGATTAATGATGACGATATGCTGCTATAGACTCACTGCTGAAT 1017  
 Db 321 IleGluMetAsnYsaIaAspAsnAspAlaGlyAsnAlaAlaIleAspSerLeuLeuAsn 340  
 Qy 1018 TATGAACCTGAGAGTATTTTAATATGAAGATATGAAGACAGATATATGATGATTT 1077  
 Db 341 TyrGlnThrValIlysrYrPheAsnAsnGlnArgYrGlnAlaGlnArgYrAspGlyPhe 360  
 Qy 1078 TTGAAGACGTATGAGACTGCTTCACTTGAAGATCACTTACTGCTGATGCTGAACCTT 1137  
 Db 361 LeuYsrThrYrGlnThrAlaSerLeuYsrThrSerThrLeuAlaMetLeuAsnPhe 380  
 Qy 1138 GGTCAAAAGTGTATTTTCAGTGTGGTTTAAACAGCTATATGTGTGCGCCAGTCAAGGA 1197  
 Db 381 GlyGlnSerAlaIlePheSerValGlyLeuThrAlaIleMetValLeuAlaSerGlnGly 400  
 Qy 1198 ATTGTGGCAGGTATCCCTTACCTGTGGAGATCTAGTAATGTGATGATGACTGCTTTTCA 1257  
 Db 401 IleValAlaGlyThrLeuThrValGlyAspLeuValMetValaGlnGlyLeuLeuPheGln 420  
 Qy 1258 CTTTCATTATCCCTGAACTTCTGGAACCTGATATATATGAGAGATGAGATCAAGCACTGATA 1317  
 Db 421 LeuSerLeuProLeuAsnPheLeuGlyThrValYrArgGlnThrArgGlnAlaLeuIle 440  
 Qy 1318 GATATGAACACTTGTCTTACTCTACAGGTAGACCCCAATTAAGAACAAGTATG 1377  
 Db 441 AspMetAsnThrLeuPheThrLeuLeuValaAspThrGlnIleYsaPlyValaThr 460  
 Qy 1378 GCATTCCTCCCTTGACATCACACCAACAGACTACGCTGGCTTGTATATGTGATTTT 1437  
 Db 461 AlaSerProLeuGlnIleThrProGlnThrAlaThrValAlaPheAspAsnValHisPhe 480  
 Qy 1438 GAATACATTGAGGCGCAAGAAAGTCTTAGTGAATATCTTTGAAGTCCCTGACAGAAAG 1497  
 Db 481 GluYrIleGluGlyGlnIlysrValaLeuSerGlyIleSerPheGlnValProAlaGlyLys 500  
 Qy 1498 AAAGTGCCATTGTAGAGGTAGTGGGTCAAGGAAACCAATATGAGGCTATTAATTT 1557

Db 501 LysValAlaIleValGlyGlySerGlySerGlyYsrSerThrIleValArgLeuLeuPhe 520  
 Qy 1558 CGCTTCTATGAGCCTCAAAAAGGATGATTTATCTTCTGGTCAAAATATATCAAGATGTG 1617  
 Db 521 ArgPheYrGlnProGlnIlysrGlySerIleYrLeuAlaGlyGlnAsnIleGlnAspVal 540  
 Qy 1618 AGCCTGAAAACCTTCGAGGGGAGTGGAGATGGTACTCAGATGCTGCTCTTCAT 1677  
 Db 541 SerLeuGlnSerLeuAspArgAlaValGlyValValaProGlnAspAlaValLeuPheHis 560  
 Qy 1678 AATTAATTTTATTAACAACCTTATATGAACAATCAAGTGTCTTCACTGAGAAAGTAT 1737  
 Db 561 AsnThrIleYrYrAsnLeuLeuYrGlyAsnIleSerAlaSerProGlnGluValYr 580  
 Qy 1738 GCAGTGCAAAATATAGCTGACCTTCATGATGCAATTCCTTGGAAATGCCACATGATGAC 1797  
 Db 581 AlaValAlaYsaLeuAlaGlyLeuHisAspAlaIleLeuArgMetProHisGlyYrAsp 600  
 Qy 1798 ACCCAAGTAGGGGAAACGAGACTCAAGCTTTCAGAGAGAGAAAAGCAAGATGACAT 1857  
 Db 601 ThrGlnValGlyGlnArgGlyLeuYsaLeuSerGlyGlyGlyYsaGlnArgValAlaIle 620  
 Qy 1858 GCAAGACCAATTTGAAAGAACCCCAAGTCAATCTATATGATGAAGTACTTCACTGTTA 1917  
 Db 621 AlaArgAlaIleLeuYsaAspProValIleLeuYrAspGlnAlaThrSerSerLeu 640  
 Qy 1918 GATTCGATTACTGAAAGACTATCTTGGTGGCCATGGAAGATGGTCCAAACACAGAACT 1977  
 Db 641 AspSerIleThrGlnGlnThrIleLeuGlyAlaMetCysAspValaIlysrAspThr 660  
 Qy 1978 TCTATTTTCATTGACACAGATTTGTCAACAGTGTGATGCAGATGAATCATTTGCTTG 2037  
 Db 661 SerIlePheIleAlaHisArgLeuSerThrValValaAspAlaAspGlnIleIleValLeu 680  
 Qy 2038 GATCAGGGTAAAGTATGCCGAACGTGGTATCCCAACAATGTTGCTTGAACCTCATAGT 2097  
 Db 681 AspGlnIlysrValaIaGlnArgGlyThrHisIleGlyLeuLeuAlaAsnProHisSer 700  
 Qy 2098 ATCTATTCAGAAATGTGGCATACAGAGACGCGTGTGAGAAACATGAATCCCAAA 2157  
 Db 701 IleYsrSerGlnMetCyrPheIsthrGlnSerSerArgValGlnAsnHisAspAsnProLys 720  
 Qy 2158 TGGGAA 2163  
 Db 721 TrpGln 722  
 RESULT 5  
 ID ABB57811  
 ID ABB57811 standard; Protein; 606 AA.  
 AC ABB57811;  
 XX 26-MAR-2002 (first entry)  
 DT 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster polyptide seq ID NO 225.  
 DE Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 KW Drosophila melanogaster.  
 OS Drosophila melanogaster.  
 XX PN MO200171042-A2.  
 XX 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US09231.  
 PF 23-MAR-2001; 2000US-191637P.  
 PR 11-UTL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA





```

QY 1438 GAATACATTAGAGGCGCAAGAAAGTCTTAGTGAATATCTTTGAAGTCCCTGACAGAAAG 1497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 356 GUTYRGUProGlyLysProIlePheArgAspLeuSerPheThrIleProIleGlyLys 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1498 AAAGTGGCCATTGTAGAGGTAGTGGGTCAAGGAAAGCAATATGTAGAGCTTATTT 1557
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 376 AsnValAlaIleValAlGlyLysSerGlyLysSerSerMetValArgLeuPhe 395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1558 CGCTCTTAGAGGCTCAAAAGGTTAGCATTTTCTTGGTGGCAAAATATCAAGATGG 1617
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 396 ArgPhePheGluProAsnSerGlyLysValLeuIleGlyGlnAspIleSerAlaVal 415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1618 AGCCTGGAAGGCTTCGAGGCGAGTGGAGTGTACTCAGAGTGTCTGCTCTTCAT 1677
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 416 AspLeuGluSerLeuArgLysValIleAlaValValProGlnAspSerValLeuPhe 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1678 AATACATTTATTAACAACCTTTATATGGAACATCATGCTTCACTGAGAGAGTAT 1737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 436 AsnThrIleGluHisAsnIleHisIleHisIleHisIleHisIleHisIleHisIle 455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1738 GAGTGGCAAAATTAAGCTGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1797
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 456 AsnAlaAlaArgMetAlaAspLeuHisAspSerIleMetSerTrpProGlyGlnTyrSer 475
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1798 ACCCAAGTAGGGAACGAGACTCAAGCTTTGAGAGGAGGAGGAGGAGGAGGAGGAGT 1857
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 ThrGlnValGlyGlnArgGlyLeuLysLeuSerGlyGlyGlnValAlaIle 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1858 GCAAGAGCCATTTTGAAGACCCCGAGTCACTATGATGATGATGATGATGATGATGATGAT 1917
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 496 AlaArgAlaIleLeuLysAsnThrProIleLeuIlePheAspGluAlaThrSerSerLeu 515
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1918 GATTGATTAAGAGAGACTATTTCTGAGTGGATGAGGAGGAGGAGGAGGAGGAGT 1977
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 516 AspSerIleThrGlnHisAsnIleLeuGlnAlaLeuThrArgAlaThrSerGlyArgThr 535
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1978 TCTATTTTCATTCACACAGATTGTCAACAGTGTGATGAGAGGAGGAGGAGGAGT 2037
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 536 SerIleCysIleAlaHisIleArgLeuSerThrValLysAspAlaAspGluIleLeuValLeu 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2038 GATCAGGTTAGGTTAGCCGAGCTGTGATCCACCATGTTGCTTCTGCTAACCCTCATAGT 2097
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 556 GUAANGlyArgValAlGlyGlnArgGlyThrHisSerGlnLeuLeu---ArgGlnAsnGly 574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2098 ACTTATTCAGGAATGTGGATACACAGACACGCCGTGTGCAGAACCATATTAACCCCAA 2157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 575 LeuTyrAlaArgLeuTrpGluThrGlnTrpGlnPhe-----AspProSer 590
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2158 TGGGAAGCAAGAAAGAAATATATATCAAGAG 2190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 591 ArgGluIleAsnGluGluValAlaAlaLysLys 601
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

PF 05-MAR-2001; 2001WO-EP02478.
XX
XX
PR 06-MAR-2000; 2000EP-0104770.
XX
XX
PA (MERE ) MERCK PATENT GMBH.
XX
XX
PI Brandt S;
XX
XX
DR WPI: 2001-589940/66.
DR N-PSDB: AAS15604.
XX
XX
PT Novel ABC transporter polypeptides for the treatment of cystic
XX fibrosis, adrenoleukodystrophy, hyperglycaemia and hypertension
XX
XX Claim 1; Page 38-39; 44pp; English.
XX
XX This sequence represents the human ABC transporter expressed in liver
XX (ATIL) protein of the invention. The invention also comprises the
XX sequence of the ATIL cDNA, an agonist or antagonist of this protein, an
XX expression system which can be used to express recombinant ATIL protein
XX or a fusion protein comprising an immunoglobulin Fc-region in a host
XX cell and an antibody immunospecific for the ATIL protein which can be
XX used to stimulate or inhibit a function. This protein may have
XX hypotensive activity and the sequences can be used in gene therapy or to
XX develop a vaccine to treat diseases, including conditions where cellular
XX resistance to toxins is induced, drug efflux is altered, intracellular
XX distribution and antigen presentation pathways are involved, such as
XX disorders related to lipid metabolism and heavy metal transport, cystic
XX fibrosis, adrenoleukodystrophy, hyperinsulinaemic hypoglycaemia and
XX hypertension. The nucleotide sequence is useful as a diagnostic reagent
XX for detecting mutations in associated genes, in chromosome localisation
XX studies, and as a valuable tool for tissue expression studies. The
XX sequences may also be useful as vaccines, and for inducing an
XX immunological responses in a mammal and are useful to configure
XX screening methods for detecting the effect of added compounds on the
XX production of mRNA and polypeptide in cells. The ATIL protein can also
XX be useful to identify membrane-bound soluble receptors.
XX
XX
SQ Sequence 842 AA;
XX
XX
XX
XX
XX Alignment Scores:
XX Pred. No.: 2,95e-107 Length: 842
XX Score: 1239.50 Matches: 286
XX Percent Similarity: 53.86% Conservative: 126
XX Best Local Similarity: 37.39% Mismatches: 240
XX Query Match: 29.14% Indels: 113
XX DB: Gaps: 16
XX
XX
XX AF133659 (1-2345) x AAU09965 (1-842)
XX
XX QY 10 CTCGGGATGCAATTTCTGGCGCTGGCGGCGGCTGTTGCAAAAGCCGCGCAC 69
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 136 LeuAlaMetGlyIleTrp-----IleLysPheArgHis 146
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 70 TCC-----GCGATTCTGATCCGGCTTTAGTCTGT 102
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 147 SerProGlyLeuLeuLeuLeuTrpThrValAlaPheAlaIleGlnValAlaVal 166
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 103 AGCGCTCAGGTCGCGCATGAGAGGCCACATCAACTCGCGGCTTGGAAACCGCTCGAGCC 162
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 167 SerTrpAsnSerProGlnTrp----- 173
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 163 TACCGATTCCAGAGTCAATTAAGATATCAATGCGAGATTGGGAAAGCAATTCA 222
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 174 -----TrpTrpAlaArgAlaAspLeuGlyGln 183
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 223 GCAGAGTCTTGAATGCTGCAAAAGGCTCTCCAGATATGCGCATGATGAAAGAGACA 282
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 184 ValGlnPhe-----SerLeuTrpValLeu----- 191
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 283 TGTGGCATGTGATGAGAGAGAGACTCCACACA----- 318
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 192 -----ArgTyrValValSerGlyLysLeuPheValLeuGlyLeuTrpAlaProGlyLeu 209
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



QY 319 -----GACCCAAAGAGGTTAAAAAGATTGATCTCGGAAA 357  
DB 210 ArgProGlnSerIYrThrLeuGlnValHlsgluluaSerGlnAspValGluArgSerGln 229  
QY 358 ATC-----ATMAAGCAATG 372  
DB 230 ValArgSerAlaAlaGlnGlnSerThrTrpArgPheGluYArgLysLeuArgLeuLeu 249  
QY 373 CTTTCTTATGTGTGGCCCAAGACAGCCAGATCTACAGCTAGAGTTGCCATTTCGCTG 432  
DB 250 SerGlyTrpLeuTrpProArgGlySerProAlaLeuGlnLeuValValLeuIleCysLeu 269  
QY 433 GGATTTTGGGTGGTGCAGAGCCATGAATATGTGTTCCCTTCAGTTTAAATATGCT 492  
DB 270 GlyLeuMetGlyLeuGluArgAlaLeuAsnValLeuValProIlePheTrpArg----- 287  
QY 493 GTACAGACCCCAACCAAGATGTCGGAAACATGCTGAACCTG-----AGTGAAGCACCA 546  
DB 288 -----AsnIleValAsnLeuLeuThrGluLysAlaPro 298  
QY 547 ---AATACAGTTGCACCATGAGCAACAGCA-----GTTCTGATGGCTAT 588  
DB 299 TrpAsnSerLeuAlaTrpThrValThrSerTrpValPheLeuLysPheLeuGlnGly 318  
QY 589 GGTGTATCAAGAGCTGAGCTGCTTTTAAACGAAGTTCCGAATGCAATTTGGCAG 648  
DB 319 GlyThrGlySerThrGly-----PheValSerAsnLeuArgThrPheLeuTrpIleArg 336  
QY 649 GTACCCCAAGATTCAATCCCAAGAAATAGCCAAATATGCTTTCATCTTCACACACTG 708  
DB 337 ValGlnGlnPheThrSerArgArgValGluLeuLeuIlePheSerHlslsLeuHlsgluLeu 356  
QY 709 GATCTGGGTTTTCACCTGAGCAGACAGACGGAGCTTATCTTAAGCTATTCAGACAGAA 768  
DB 357 SerLeuArgTrpHlslsLeuGlyArgArgThrGlyGluValLeuAlaGlyIleAlaAspArgGly 376  
QY 769 ACAAGGGGTACAGTTTGTCTGAGTGCCTTGGTATTTAATCTTCCCATCATGCTTT 828  
DB 377 ThrSerSerValThrGlyLeuLeuSerTrpLeuValPheAsnValIleProThrLeuAla 396  
QY 829 GAATGATGCTGTGCAGTGGTCTTTGATATACAAATGC-----GGTCCCACTTTGCT 882  
DB 397 Aspllellelle-----GlyIlellelleTrpPheSerMetPhePheAsnAlaTrpPheGly 414  
QY 883 TTGTAACCTTGGAAACCTGTATACACAGACTTCACAGTTGCAGTGCAGTGCAGTGC 942  
DB 415 LeuIleValPheLeuCysMetSerLeuTrpLeuThrIleValIleValThrGluTrp 434  
QY 943 AGAATGATTTAGAAATGAATGAACAAAGCAGATATGACAGTAATGCTGTATA 1002  
DB 435 ArgThrLysPheArgGArgAlaMetAsnThrGlnLysAsnHlslsAlaArgAlaVal 454  
QY 1003 GACTCATCTGTAATTTGAACCTGTGAAGTATTTAATGAAGAAATGAAGACACAG 1062  
DB 455 AspSerLeuLeuAsnPheGluThrValLysTrpTrpAsnAlaGlnSerTrpGluValGlu 474  
QY 1063 AGATATGATGATTTTGAAGACGTATGAGACCTGCTTCAATGAAGATACCTTACTG 1122  
DB 475 ArgTrpArgGluAlaIlellelleLysTrpGlnLysLeuGluTrpLysSerSerAlaSerLeu 494  
QY 1123 GCTATGCTGAACCTTGGTGAAGTGTATTTTCAGTTCGCTTAAACAGATTAATGCTG 1182  
DB 495 ValLeuLeuAsnGlnThrGlnAsnLeuValIleGlyLeuGlyLeuLeuAlaGlySerLeu 514  
QY 1183 CTGCCCAGCTAGAGGAATTTGGCAGGTACCTTACTGTTGAGATCTAGTATGCTGAAT 1242  
DB 515 LeuCysAlaTrpPheValThrGluGlnLysLeuValGlyAspTrpValLeuPheGly 534  
QY 1243 GGAATGCTTTTTCAGCTTCAATTCACCCCTGAACCTTCTGGAACTGATATTAAGAGACT 1302  
DB 535 ThrTrpIleIleGlnLeuTrpMetProLeuAsnTrpPheGlyTrpTrpArgMetIle 554

QY 1303 AGACACAGACTCATAGATATGAACACCTTGTATTACTTACCAAGGTAGACACCAATT 1362  
DB 555 GlnThrAsnPheIleAspMetGluAsnMetPheAsnLeuLysGluGlnThrGluLys 574  
QY 1363 AAAGACAAAGATGATGATCTCCCTTCAGATCAACACAGACAGATCAACCGTGGCCTT 1422  
DB 575 LysAspLeuProGlyAlaGlyProLeuArgPhe-----GlnLysGlyArgIleGluPhe 592  
QY 1423 GATTAATGCTATTTGAATCATTTAGAGGCGCAGAAAGCTCTTACTGAAATATCCTTTGAA 1482  
DB 593 GluAsnValHlslsPheSerTrpAlaAspGlyArgGluThrLeuGlnAspValSerPheThr 612  
QY 1483 GTCCCTCAGAGAAAGAAAGTGGCCATTGTAGAGGTAGTGGGTGAGGAAAGCAATA 1542  
DB 613 ValMetProGlyGlnThrLeuAlaLeuValGlyProSerGlyAlaGlyLysSerThrIle 632  
QY 1543 GTAGAGCTATTTATTTGCTTCTTATGACCCCAAAAGGTAGCAATTTATCTTGCTGTCAA 1602  
DB 633 LeuArgLeuLeuPheArgPheTrpAspIleSerSerGlyCysIleArgIleAspGlyGln 652  
QY 1603 AATATCAAGATGTAGAGCCCTGGAAGACCTTCGAGAGGCGAGTGGAGTGAACCTCAGAGT 1662  
DB 653 AsplleSerGlnValThrGlnAlaSerLeuArgSerHlslsGlyValValProGlnAsp 672  
QY 1663 GCTGCTCTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1722  
DB 673 ThrValLeuPheAsnAspThrIleAlaAspAsnIleArgTrpGlyArgValThrAlaGly 692  
QY 1723 CTTGAGAAATGTATGAGAGTGGCAAAATTTAGCTGAGCTTCAATGATGAAATTTCTTGAATG 1782  
DB 693 AsnAspGluValGluAlaAlaGlnAlaAlaGlyIleHlslsAspAlaIleMetAlaPhe 712  
QY 1783 CCAATGATATGACACCCCAAGTGGGGAACGAGACTCAAGCTTTCAGAGGAGAAAG 1842  
DB 713 ProGlnGlyTrpArgThrGlnValGlyGluArgGlyLeuLysSerGlyGlyGluLys 732  
QY 1843 CAAGAGTACCAATTCAGACAGCCATTTTGAAGACCCCAAGTCAATCTTATGATGAA 1902  
DB 733 GlnArgValAlaIleAlaIleAlaArgThrIleLeuLysAlaProGlyIleIleLeuLysGlu 752  
QY 1903 GCTAATTCATCGTTAATTTGATTTACTGAGAGACATTTCTTGGTGCATGAAAGATGTG 1962  
DB 753 AlaThrSerAlaLeuAspThrSerAsnGluArgAlaIleGlnAlaSerLeuAlaLysVal 772  
QY 1963 GTCAAACACAGACTTCTATTTTCAATTCAGACAGATTTGCAACAGGTGTGATGCAAT 2022  
DB 773 CysAlaAsnArgThrThrIleValAlaAlaHlslsArgLeuSerThrValAlaAsnAlaAsp 792  
QY 2023 GAAATCATTTGCTTGGATCAGGTAAAGTAAAGCCGAAAGTGTACCCATGCTTTCCTT 2082  
DB 793 GlnIleLeuValIleLysAspGlyCysIleValGluArgGlyArgHlslsAlaLeuLeu 812  
QY 2083 GCTAACCTCATATGATCTTATTTCAAAATGTGGCATATACAGACAGACCGCTGACAGAC 2142  
DB 813 SerArg---GlyGlyValTrpAlaAspMetTrpGlnLeuGlnGlnGlnGluThr 831  
QY 2143 CATGATTAACCCCAA 2157  
DB 832 SerGluAspThrLys 836

RESULT 8  
AAE16764  
ID AAE16764 standard; Protein; 842 AA.  
XX  
XX  
AAE16764;  
XX  
XX 09-APR-2002 (first entry)  
DE Human transporter and ion channel-1 (TRICH-1) protein.  
XX  
XX Human; transporter and ion channel-1; TRICH-1; neuroprotective; asthma;  
KW nootropic; cytosolic; cardiovascular; immunosuppressive; cardiomyopathy;  
KW antiinflammatory; protein therapy; akinesia; cystic fibrosis; leukaemia;



KM Bell's palsy; amyotrophic lateral sclerosis; Alzheimer's disease; cancer;  
 KM amnesia; dementia; myocarditis; Duchenne's muscular dystrophy; AIDS;  
 KM Acquired Immune Deficiency Syndrome; Addison's disease; allergy; angina;  
 KM cell proliferative disorder; psoriasis; cardiac disease; hypertension;  
 KM bradyarrhythmia; gene expression; drug screening.

XX Homo sapiens.

XX Key location/Qualifiers

XX Domain 186..203 /label= Transmembrane\_domain

XX Domain 386..406 /label= Transmembrane\_domain

XX WO200192304-A2.

XX 06-DEC-2001.

XX 25-MAY-2001; 2001WO-US17065.

XX 26-MAY-2000; 2000US-208424P.

XX 01-JUN-2000; 2000US-209001P.

XX 08-JUN-2000; 2000US-210588P.

XX 16-JUN-2000; 2000US-212335P.

XX 22-JUN-2000; 2000US-213747P.

XX 29-JUN-2000; 2000US-215391P.

XX (INCY-) INCYTE GENOMICS INC.

XX Thornton M, Walla NK, Yue H, Nguyen DB, Lai P, Gandhi AR;

XX Trilbouley CM, Yao MG, Ramkumar J, Au-Young J, Lu Y, Tang YF;

XX Azimzai Y, Bruns CM, Griffin JA, Yang J, Sanjanwala MS;

XX Rammann BE, Lee EA, Hafalia A, Baughn MR, Green BD, Khan PA;

XX Kearney L, Elliott VS, Sellhammer JT, Policky JL, Borowsky ML;

XX Burford N, Ding L, Lu DM, Hillman JL;

XX WPI; 2002-122055/16.

XX N-PSDB; AAD27254.

XX New human transporters and ion channels (TRICH) polypeptides useful for

XX diagnosing, treating or preventing disorders associated with aberrant

XX expression of TRICH -

XX Claim 1; Page 144-146; 210pp; English.

XX The invention relates to human transporters and ion channels (TRICH)

XX polypeptides and their cDNA molecules. The nucleic acid and polypeptide

XX sequences are useful in the diagnosis, treatment, and prevention of

XX disorders associated with transport (akinesia, cystic fibrosis, Bell's

XX palsy, amyotrophic lateral sclerosis); neurological (Alzheimer's disease,

XX amnesia, dementia); muscle (cardiomyopathy, myocarditis, Duchenne's

XX muscular dystrophy); immunological (AIDS, Addison's disease, allergies,

XX asthma); cell proliferative disorders (cancers, leukaemia, psoriasis);

XX cardiac disease (angina, hypertension, or bradyarrhythmia) and in the

XX assessment of the effects of exogenous compounds on the expression of

XX nucleic acid and amino acid sequences of transporters and ion channels.

XX The polynucleotides may be used to detect and quantify gene expression

XX in biopsied tissues in which TRICH expression may be correlated with a

XX disease, to generate hybridization probes for mapping naturally occurring

XX genomic sequence, and in drug screening. The present sequence is human

XX TRICH-1 protein.

XX Sequence 842 AA;

XX Alignment Scores:

XX Pred. No.: 2,956-107 Length: 842

XX Score: 1239,50 Matches: 286

XX Percent Similarity: 53,86% Conservative: 126

XX Best Local Similarity: 37,39% Mismatches: 240

XX Query Match: 29,14% Indels: 113

XX DB: 23 Gaps: 16

AF133659 (1-2345) x AAE16764 (1-842)

QY 10 CTCGCGATGCAATTCCTGGCGCGCGCGCGCTGCTTGAAGGCGCGCAC 69  
 |||||  
 DB 136 LeuAlaMetGlyIleTrp-----IleValPheArgHis 146  
 QY 70 TCC-----CGAATTCGATCCGCGCTTATGCTCTGTT 102  
 |||||  
 DB 147 SerProGlyLeuLeuLeuTrpThrValAlaPheAlaIleVal 166  
 QY 103 AGCGGCTCAGGTCGCGAGTGAGGCGACATCACTCGGCGCTTGGAAACGCTCGAGCC 162  
 |||||  
 DB 167 SerTrpAsnSerProGlnTrp----- 173  
 QY 163 TACCAGATTCAGAGATCTTAAAAAGTATCATGCGAGATTGGGAAAGGCAATTCA 222  
 |||||  
 DB 174 -----TrpThrAlaArgAlaAspLeuGlyGln 183  
 QY 223 GGACAGTCTTATGATGCTGCAAGAGCTCTCCAGGTATGCGCATGTATGAAAGAGACA 282  
 |||||  
 DB 184 ValGlnPhe-----SerLeuTrpValLeu----- 191  
 QY 283 TGTGGCATGTCATGACGAGAGACTCCACACA----- 318  
 |||||  
 DB 192 -----ArgTrpValValSerGlyGlyLeuPheValLeuGlyLeuTrpAlaProGlyLeu 209  
 QY 319 -----GACCCAAAAGAGAGGTTAAAGATGTTGATCTCGCAA 357  
 |||||  
 DB 210 ArgProGlnSerTrpThrLeuGlnValHisGlnGluAspGlnAspValGluArgSerGln 229  
 QY 358 ATC-----ATAAAGCAATG 372  
 |||||  
 DB 220 ValArgSerAlaAlaGlnGlnSerThrTrpArgAspPheGlyArgValLeuArgLeuLeu 249  
 QY 373 CTTTCTTATGTGTGGCCCAAGACAGCAGATCTAGACAGTACAGTTGCCATTTGCTG 432  
 |||||  
 DB 250 SerGlyTrpLeuTrpProArgGlySerProAlaLeuGlnLeuValValLeuIleCysLeu 269  
 QY 433 GGATTTTGGGTGGTGCAGAAAGCCATGATATTTGTGTTCCCTTCATGTTAAATATGCT 492  
 |||||  
 DB 270 GlyLeuMetGlyLeuGluArgAlaLeuAsnValLeuValProIlePheTrpArg----- 287  
 QY 493 GTAGACAGCTCAACAGATGTCGGGAAACATGCTGAACCTG-----AGTAGTCACCA 546  
 |||||  
 DB 288 -----AsnIleValAsnLeuThrGlnValAlaPro 298  
 QY 547 ---AATACAGTTCACACATGCGCAACACA-----GTTTCGATGGCGTAT 588  
 |||||  
 DB 299 TrpAsnSerLeuAlaTrpThrValThrSerTrpValPheLeuValPheLeuGlnGlyGly 318  
 QY 589 GGTGATACAGAGCTGAGCTGCTTTTAAACAGAGTTCGAAATGCAATTTGGCAAG 648  
 |||||  
 DB 319 GlyThrGlySerThrGly-----PheValSerAsnLeuAlaGlnPheLeuTrpIleArg 336  
 QY 649 GTAGCCAGAAATTCATATCGAAGAAATAGCCAAATATGCTTTCCATCTTCACAACTG 708  
 |||||  
 DB 337 ValGlnGlnPheThrSerArgValGluLeuLeuIlePheSerHisLeuHisGluLeu 356  
 QY 709 GATCTGGGTTTTCACCTGAGACAGACAGCGGACCTTATATTAAGCGTATGACAGGCA 768  
 |||||  
 DB 357 SerLeuAlaGlyTrpHisLeuGlyAlaArgThrGlyGluValLeuAspGlyIleValAspArgGly 376  
 QY 769 ACAAGGGGTATCAGTTTGTCTGAGTCTTGTGATTTAATCTTCCTCCATCAATGTT 828  
 |||||  
 DB 377 ThrSerSerValThrGlyLeuLeuSerTrpLeuValPheAsnValIleProThrLeuAla 396  
 QY 829 GAAGTATGCTTGTACAGTGTGTGTTGTATTAACAATGC-----GGTCCCAAGTTGCT 882  
 |||||  
 DB 397 AspIleIleIle-----GlyIleIleTrpPheSerMetPhePheAsnAlaTrpPheGly 414  
 QY 883 TTGGTAACTTGTGAACACTGTGTACATACACAGATTCACAGTTGACGTACACAGCGTG 942  
 |||||  
 DB 415 LeuIleValPheLeuCysMetSerLeuTrpLeuThrIleValIleValThrGluTrp 434



The sequence represents a human ATP binding cassette (ABC) transporter family member, ABCB12. The nucleic acid encoding ABCB12 is useful as a target for developing modulating agents of multidrug resistance and as diagnostic and therapeutic tools to control cellular transport in cells of the brain and transport across the blood-brain-barrier and for developing strategies to deliver drugs to the brain, for treating Alzheimer's disease, hyperinsulinaemic hypoglycaemia, cystic fibrosis, atherosclerosis, macular degeneration, adrenoleukodystrophy, prion diseases, Huntington's disease, Parkinson's disease and for treating mood and panic disorders. The nucleic acid is useful for expressing ABCB12, to detect ABCB12 transporter mRNA or genetic alteration in an ABCB12 transporter gene, and to modulate ABCB12 transporter activity and is useful in tissue-typing to identify individuals in forensic biology. The protein is useful as an immunogen to raise anti-ABCB12 transporter antibodies and is used to screen for naturally occurring ABCB12 transporter substrates and to screen for drugs or compounds which modulate ABCB12 transporter activity. The nucleic acid and protein are used to treat a subject having a disorder characterised by aberrant or unwanted ABCB12 transporter protein or nucleic acid expression or activity. The nucleic acid, protein and antibodies against ABCB12 are useful in screening assays, predictive medicine and in prophylactic and therapeutic treatment. The antibodies are useful to detect and isolate ABCB12, regulate the bioavailability of ABCB12 and to modulate ABCB12 transporter activity. A vector expressing the nucleic acid is useful for producing non-human transgenic animals.

XX Sequence 843 AA:

Alignment Scores:

|                        |           |               |     |
|------------------------|-----------|---------------|-----|
| Pred. No.:             | 1,85e-106 | Length:       | 843 |
| Score:                 | 1231.00   | Matches:      | 284 |
| Percent Similarity:    | 53.57%    | Conservative: | 129 |
| Best Local Similarity: | 36.84%    | Mismatches:   | 250 |
| Query Match:           | 28.94%    | Indels:       | 108 |
| DB:                    | 22        | Gaps:         | 16  |

AF133659 (1-2345) x AAU00010 (1-843)

```

QY 10 CTCGGATGTCATTTCTGGCGCTGCGCGCGCGCTTTCGAAAGCCGCGCAC 69
DB 136 LeuAlaMetGlyIleTrp-----IleYsPhehRghIs 146
QY 70 TCC-----GCGATTCTGATCCGCGCTTATGCTCTGT 102
DB 147 SerProGlyLeuLeuLeuLeuTrpThrValAlaIlePheAlaIleGluLeuVal 166
QY 103 AGCGGCTCAGTCCGCGAGTGGAGGCCAATCGCGCTTGAGAAACGCTCGAGCC 162
DB 167 SerTyrPheSerProGlnTrp----- 173
QY 163 TACCAGATTCCAGAGTCATTAAAGTATCACATGCGCAGATTTGGGAAGCAATTCA 222
DB 174 -----TrrTPAlaAlaArgAlaAspLeuIleGln 183
QY 223 GGACAGTTCTTAGATGCTGCAAAAGGCTCTCCAGTATGCGCCTATGAAAGAGACA 282
DB 184 ValGlnPhe-----SerLeuTrpValLeu----- 191
QY 283 TGTTCGATGTCATGACGAGAGAGACTCCACACA----- 318
DB 192 -----ArgTyrValValSerGlyIleuPheValLeuGlyLeuTrpAlaProGlyLeu 209
QY 319 -----GACCCAAAGAAAGGTTAAAGATCTTATCTCGGAAA 357
DB 210 ArgProGlnSerTyrThrLeuGlnValHisGlnGluAspValAlaGluArgSerGln 229
QY 358 ATC-----ATTAAGCAATG 372
DB 230 ValArgSerAlaAlaGlnIleSerThrTrpArgAspPheGlyArgLysLeuArgLeuLeu 249
QY 373 CTTTCTTATGTTGGCCCAAGACAGCCAGATCTACGAGCTAGAGTTGCCATTGCGCTG 432
DB 250 SerGlyTyrLeuTrpProArgLysSerProAlaLeuGlnLeuValValLeuIleCysLeu 269

```

```

QY 433 GAATTTTGGGNGCAAGGCCATGATATTTGGTTCCCTTCATGTTTAAATATGCT 492
DB 270 GlyLeuMetGlyLeuGluArgAlaLeuAsnValLeuValProIlePheTyrArgAspIle 289
QY 493 GTAGACACCTTCAACAGATATCGGGAAACATGCTGAACCTGAGTATGACCA--AAT 549
DB 290 ValAsnLeuLeuThrGlu-----LysAlaProTyrAsn 300
QY 550 ACAGTTGCA---ACCATGGCAACAGACTTCTGATGCTAT-----GGGTATCAAGA 600
DB 301 SerLeuAlaTrpThrValThrThrTyrValaPheLeuLysPheLeuGlnGlyGlyThr 320
QY 601 GCTGGAGCTGCTTTTTCATCAAGAAATTGCAATGACGATTTGGCAAGTGGCCAGAAAT 660
DB 321 GlySerThrGlyPheValSerAsnLeuArgThrPheLeuTrpIleArgValGlnGlnPhe 340
QY 661 TCATTCGAGAAATAGCCAAAAATGTCTTCTCCATCTTCACAACTGGATCTGGGTTT 720
DB 341 ThrSerArgArgValGluLeuLeuIlePheSerHisLeuHisGluLeuSerLeuArgTyr 360
QY 721 CACCTGACACAGACACGGGAGCTTTATCTAAGCTATTTACAGAGAACAAAGGGTATC 780
DB 361 HisLeuGlyArgArgThrGlyGluValLeuArgIleAlaAspArgGlyThrSerSerVal 380
QY 781 AGTTTGTCTGAGTGTGTTGGTATTTATCTTCTCCCATGCTTTGAAGTATGCTT 840
DB 381 ThrGlyLeuLeuSerTyrLeuValaPheAsnValIleProThrLeuAlaAspIleIle 400
QY 841 GTCACTGCTGTTTGTATTAACAATGC-----GGTCCCACTTGTCTTGGTAAACCTT 894
DB 401 -----GlyIleIleTyrPheSerMetPhePheAsnAlaTrpPheGlyLeuIleValaPhe 418
QY 895 GGAACACTGTGTACATACACAGCATTCACACTTCAGTTCACAGCTGAGAGAACTAGATT 954
DB 419 LeuCysMetSerLeuTyrLeuThrLeuThrIleValaValThrGluTrpArgThrIysPhe 438
QY 955 AGAATGAATGAATGAACAAAGCAATATGATGACAGTATGCTGCTATGACTCATCTGCT 1014
DB 439 ArgArgAlaMetAsnThrGlnGluAsnAlaThrArgAlaArgAlaValaAspSerLeuLeu 458
QY 1015 AATTATGAACCTGTGAGATATTTAATATGAAGATATGAAGACACAGATATATGATGA 1074
DB 459 AsnPheGluThrValValTyrTyrAsnAlaGluSerTyrGluValaGluArgTyrArgGlu 478
QY 1075 TTTTGAAGAGTATGACACTGCTTATGAAAGTACCTCTATCTGCTATGCTGTAAC 1134
DB 479 AlaIleIleLysTyrGlnGlyLeuGluTrpLysSerSerAlaSerLeuValLeuLeuAsn 498
QY 1135 TTGTGTCAAAGTCAATTTTCAGTGTCCGTTTAAACAGCTATATATGCTCGCCAGTCA 1194
DB 499 GlnThrGlnAsnLeuValIleGlyLeuGlyLeuLeuAlaGlySerLeuLeuCysAlaIleTyr 518
QY 1195 GGAAATGTGGCAGGTACCTTACTGTTGAGATCTAGTAAATGTGGTAAATGACTGCTTTT 1254
DB 519 PheValThrGlnGlnIleLeuGlnValaGlyAspTyrValaLeuPheGlyThrTyrIleIle 538
QY 1255 CAGCTTCAATTAACCTTGAACCTTCTGGGAACCTGATATTAAGACACTAGCAACGACTC 1314
DB 539 GlnLeuTyrMetProLeuAsnTrpPheGlyThrTyrArgMetIleGlnThrAsnPhe 558
QY 1315 ATAGATATGAACACCTTGTATCTACTTCAAGAGTACACCCAAATTAAGACAAAGTG 1374
DB 559 IleAspMetGluAsnMetPheAspLeuLeuLysGluTrpGluValaLysAspLeuPro 578
QY 1375 ATGGATCTCCCTTCAGATCAACACAGACAGCTACCGTGGCTTTGATATATGTCAT 1434
DB 579 GlyAlaGlyProLeuArgPhe-----GlnLysGlyArgGluIleGluPheGluAsnValHis 596
QY 1435 TTTGAATACATTGAAGGCCAAGAAAGTCTTATGATGATATCTTTGAAAGTCCCTGCACGA 1494
DB 597 PheSerTyrAlaAspGlyArgGluThrLeuGlnAspValSerPheThrValaMetProGly 616

```

|           |                 |   |      |
|-----------|-----------------|---|------|
| QY        | 1495            | AMGAAGAGCCGACATTGTACGAGTGTGGGTGACGGGAAAGACAATACGAGGCGATTA             | 1554 |
| Db        | 617             | glnthrleualaleuvalajlyproserglyalaaglylysetthrilleuvalrgleu           | 636  |
| QY        | 1555            | TTTCGCTTCATGAGCGCTTCAAAAAGGTACATTTATCTTCGTGGTCAAAATATACAGAT           | 1614 |
| Db        | 637             | PheArgPheTyAspIleSerSerGlyCybIleArgIleAspGlyGlnAspIleSerGln           | 656  |
| QY        | 1615            | GTGAGCGCTGGAAAGCCTTGGAGGGCAGTGGAGGTGACTTCAGAGTCTGCTCTTC               | 1674 |
| Db        | 657             | ValThrGlnAlaSerLeuArgSerHisIleGlyValValProGlnAspThrValLeuPhe          | 676  |
| QY        | 1675            | CATTATACATTTATTAACAACCTCTTATATGGAACATGACAGGCTTCACCTAGAGAGG            | 1734 |
| Db        | 677             | AsnAspThrIleAlaGlnAsnIleArgTyrGlyArgValThrAlaGlyAsnAspGluVal          | 696  |
| QY        | 1735            | TATGACAGTGGCAAAATTAGCTGGACTTCATGATGCATTTCTTCGATCGACATGGATAT           | 1794 |
| Db        | 697             | GluAlaAlaAlaGlnAlaAlaGlyIleHisAspAlaIleMetAlaPheProGlnGlyTyr          | 716  |
| QY        | 1795            | GACACCCAGTAGGGGGAACGAGACTCAAGCTTTCAGAGAGAAAGCAAGAGTAGCA               | 1854 |
| Db        | 717             | ArgThrGlnValGlyGlnArgGlyLeuLeuSerGlyGlyGlnLysGlnProValAla             | 736  |
| QY        | 1855            | ATTGCAAGAGCCATTTTGAAAGACCCCGACATCTTCTATGATGAGACTTACTTCATCG            | 1914 |
| Db        | 737             | IleAlaArgThrIleLeuLysAlaProGlyIleIleLeuLeuAspGlnAlaThrSerAla          | 756  |
| QY        | 1915            | TTAGATTGGATTACGGAAGAGACTTCTTCGTGGCCATGAAGAGTGTGTCAAACACAGA            | 1974 |
| Db        | 757             | LeuAspThrSerAsnGlnLysAlaIleGlnAlaSerLeuAlaLysValCysAlaAsnArg          | 776  |
| QY        | 1975            | ACTCTATTTTTCATGACACACAGATTTGTCAACAGTGTGATGACAGATGAATCTATTC            | 2034 |
| Db        | 777             | ThrThrIleValValAlaHisArgLeuSerThrValValAsnAlaAspGlnIleLeuVal          | 796  |
| QY        | 2035            | TTGGAATCAGGGTAAAGTAGCCGCAAGCTGTATCCACATGCTGTTGCTTGCTAACCTCAT          | 2094 |
| Db        | 797             | IleLysAspGlyCysAlaIleValGlnLysArgIleArgHisGlnAlaLeuLeuSerArg--Gly     | 815  |
| QY        | 2095            | AGTACTATTTTCGAATGTGGCGCATACACAGACACCGCGTGCAGAACCATGATAACCC            | 2154 |
| Db        | 816             | GlyValTyrAlaAspMetCysIleLeuGlnGln-----GlnGlyGlnGlnGlnThr              | 832  |
| QY        | 2155            | AAATGGGAAGCAAGAAAGAAATATATTCACA                                       | 2187 |
| Db        | 833             | SerGluAspThrLysProGlnThrMetGluArg                                     | 843  |
| RESULT 10 |                 |   |      |
| AAU00011  | ID              | AAU00011 standard; Protein; 843 AA.                                   |      |
| XX        | AC              | AAU00011;   |      |
| XX        | DT              | 08-MAY-2001 (first entry)   |      |
| XX        | DE              | Human ABCB12 transporter.   |      |
| XX        | KM              | Human; ABCB12 transporter; ATP binding cassette; Alzheimer's disease; |      |
| KM        | KW              | hyperinsulinaemic hypoglycaemia; cystic fibrosis; atherosclerosis;    |      |
| KM        | KW              | macular degeneration; adrenoleukodystrophy; prion disease;            |      |
| KW        | KW              | Huntington's disease; Parkinson's disease; mood disorder; panic;      |      |
| XX        | XX              | antibody; immunogen.  |      |
| OS        | XX              | Homo sapiens.   |      |
| XX        | XX              |   |      |
| FT        | Key             | Location/Qualifiers   |      |
| FT        | Domain          | 27..50  |      |
| FT        |                 | /label= Transmembrane_helix_1   |      |
| FT        | Misc-difference | 254   |      |
| FT        |                 | /note= "Wild-type Trp replaced with Xaa. Encoded by                   |      |
| FT        |                 | TGG. This substitution is probably due to a misprint in               |      |
| FT        |                 | the specification"  |      |

| FT                  | Domain  | 262..285          |
|---------------------|---|-------------------|
| FT                  | /label= Transmembrane_helix_2   |                   |
| FT                  | Domain  | 304..326          |
| FT                  | /label= Transmembrane_helix_3   |                   |
| FT                  | Domain  | 383..405          |
| FT                  | /label= Transmembrane_helix_4   |                   |
| FT                  | Domain  | 503..521          |
| FT                  | /label= Transmembrane_helix_5   |                   |
| FT                  | Domain  | 530..548          |
| FT                  | /label= Transmembrane_helix_6   |                   |
| FT                  | Region  | 623..630          |
| FT                  | /label= Walker_A_motif  |                   |
| FT                  | Region  | 727..738          |
| FT                  | /label= ATP-binding_cassette_signature                                    |                   |
| FT                  | Region  | 739..754          |
| FT                  | /label= Walker_B_motif  |                   |
| XX                  | WM000114547-A1.   |                   |
| XX                  | 01-MAR-2001.  |                   |
| XX                  | 18-AUG-2000; 2000MO-CA00964.  |                   |
| XX                  | 20-AUG-1999; 99US-0161724.  |                   |
| XX                  | 30-AUG-1999; 99US-0151473.  |                   |
| XX                  | 17-AUG-2000; 2000US-0641353.  |                   |
| PA                  | (ACT1-) ACTIVEPASS PHARM INC.   |                   |
| PL                  | Le Bihan S, Wilson C;   |                   |
| XX                  | WPI; 2001-218448/22.  |                   |
| DR                  | N-PSDB; AAS00010.   |                   |
| XX                  |   |                   |
| PT                  | Novel isolated ABCB12 transporter nucleic acid molecule useful as         |                   |
| PT                  | target for developing modulating agents of multidrug resistance, and in   |                   |
| PT                  | diagnosis and therapy for Alzheimer's disease and, mood and panic         |                   |
| PT                  | disorders -   |                   |
| XX                  |   |                   |
| PS                  | Claim 12; Fig 2; 97pp; English.   |                   |
| XX                  |   |                   |
| CC                  | The sequence represents a human ATP binding cassette (ABC) transporter    |                   |
| CC                  | family member, ABCB12. The nucleic acid encoding ABCB12 is useful as a    |                   |
| CC                  | target for developing modulating agents of multidrug resistance and as    |                   |
| CC                  | diagnostic and therapeutic tools to control cellular transport in cells   |                   |
| CC                  | of the brain and transport across the blood-brain-barrier and for         |                   |
| CC                  | developing strategies to deliver drugs to the brain, for treating         |                   |
| CC                  | Alzheimer's disease, hyperhomocinaemic hypoglycaemia, cystic fibrosis,    |                   |
| CC                  | atherosclerosis, macular degeneration, adrenoleukodystrophy, prion        |                   |
| CC                  | diseases, Huntington's disease, Parkinson's disease and for treating      |                   |
| CC                  | mood and panic disorders. The nucleic acid is useful for expressing       |                   |
| CC                  | ABCB12, to detect ABCB12 transporter mRNA or genetic alteration in an     |                   |
| CC                  | ABCB12 transporter gene, and to modulate ABCB12 transporter activity and  |                   |
| CC                  | is useful in tissue-typing to identify individuals in forensic            |                   |
| CC                  | biology. The protein is useful as an immunogen to raise anti-ABCB12       |                   |
| CC                  | transporter antibodies and is used to screen for naturally occurring      |                   |
| CC                  | ABCB12 transporter substrates and to screen for drugs or compounds which  |                   |
| CC                  | modulate ABCB12 transport activity. The nucleic acid and protein are used |                   |
| CC                  | to treat a subject having a disorder characterised by aberrant or         |                   |
| CC                  | unwanted ABCB12 transporter protein or nucleic acid expression or         |                   |
| CC                  | activity. The nucleic acid, protein and antibodies against ABCB12 are     |                   |
| CC                  | useful in screening assays, predictive medicine and in prophylactic and   |                   |
| CC                  | therapeutic treatment. The antibodies are useful to detect and isolate    |                   |
| CC                  | ABCB12, regulate the bioavailability of ABCB12 and to modulate ABCB12     |                   |
| CC                  | transporter activity. A vector expressing the nucleic acid is useful for  |                   |
| CC                  | producing non-human transgenic animals.                                   |                   |
| XX                  |   |                   |
| SO                  | Sequence 843 AA;  |                   |
| Alignment Scores:   |   |                   |
| Pred. No.:          | 3..08e-105  | 843               |
| Score:              | 1218..00  | 283               |
| Percent Similarity: | 53.44%  | Conservative: 129 |

Best Local Similarity: 36.71% Mismatches: 251  
Query Match: 28.63% Indels: 108  
DB: 22 Gaps: 16  
AF133659 (1-2345) x AAU00011 (1-843)

QY 10 CTCGGAGTCATCTTGGCGCGCGCGCGCGCTGCTTGGAAAAAGCCGGCAC 69  
DB 136 LeuAlaMetGly1LeuTrp-----1LeuSphaArgHis 146  
QY 70 TCC-----GGGATTCGATCCGGCCTTACTCTCTT 102  
DB 147 SerProGlyLeuLeuLeuLeuTrpThrValAlaPheAlaAlaGluAsnLeuAlaLeuVal 166  
QY 103 AGCGGTCAGATCCGCAAGTGGAGGCCACATCAACTCGGCGCTGGGAAACGCTCGAGCC 162  
DB 167 SerTrpAsnSerProGlnTrp----- 173  
QY 163 TACCAAGATTCACAGATCATTTAAAAAGTATCAGATGGCAGAGATTGGAAAAAGCAATTCA 222  
DB 174 -----TTrpAlaArgAlaAspLeuGlyGlnGln 183  
QY 223 GGACAGTTCTTAGATGCTGCAAAAGGCTCTCCAGGTATGGCCATGATAGAAAAAGAGACA 282  
DB 184 ValGlnPhe-----SerLeuTrpValLeu----- 191  
QY 283 TGTGGCATGTCATGACGAGAGAGAGACTCCACACA----- 318  
DB 192 -----ArgTyrValValSerGlyLeuPheValLeuGlyLeuTrpAlaProGlyLeu 209  
QY 319 -----GACCCAAAAAGAGGTTAAAAAGATGTGATACCTGGAAA 357  
DB 210 ArgProGlnSerTyrThrLeuGlnValHisGluGlnAspValGluArgSerGln 229  
QY 358 ATC-----ATAAAACGATNG 372  
DB 230 ValArgSerAlaAlaGlnGlnSerThrTrpArgAspPheGlyArgValLeuArgLeuLeu 249  
QY 373 CTTTCTTATGTTGGCCCAAGACAGCCAGATCTACAGATGAGATTGCCATTTCCGCTG 432  
DB 250 SerGlyTyrLeu\*\*\*ProArgGlySerProAlaLeuGlnLeuValValLeuLeuLeuLeu 269  
QY 433 GGATTTTGGGTGGTGCAAAAGCCATGAATATTTGGTTCCTTCATGTTTAAATATGCT 492  
DB 270 GlyLeuMetGlyLeuGlnLysArgAlaLeuAsnValLeuValProIlePheTyrArgAspIle 289  
QY 493 GTAGACAGCTCAACCAAGTGTCCGAAAAATGCTGAACCTGAGATGACCAACA---AAT 549  
DB 290 ValAsnLeuLeuThrGlu-----LysAlaIleProTrpAsn 300  
QY 550 ACAGTTGCA---ACCATGGCAACAGAGAGTTCTGATTTGGCTAT-----GGTGTATCAGA 600  
DB 301 SerLeuAlaIleTrpThrValThrThrTyrValPheLeuLysPheLeuGlnGlyGlyThr 320  
QY 601 GCTGAGAGCTCTTTTAAACGAAGTTGAAATGCAATTTTGGCAGAGTACCCAGAAAT 660  
DB 321 GlySerThrGlyPheValSerAsnLeuArgThrPheLeuTrpIleArgValGlnGlnPhe 340  
QY 661 TCATATCCGAAGAATAGCCAAAATGCTTTCTCCATCTTCAACACTGGATTTGGGTTT 720  
DB 341 ThrSerArgValGluLeuLeuIlePheSerHisLeuHisGluLeuSerLeuArgTrp 360  
QY 721 CACCTGAGACAGACAGCGGAGCTTATCTAAGCTATTGACAGAGAAACAGAGGATAC 780  
DB 361 HisLeuGlyArgArgThrGlyValLeuArgGlyLeuAlaAspArgGlyThrSerSerVal 380  
QY 781 AGTTTGTCTGAGTGTCTTGTATTAATCTTCCCATCATGTTTGAAGTATGCTT 840  
DB 381 ThrGlyLeuLeuSerTyrLeuValPheAsnValIleProThrLeuAlaAspIleIle 400  
QY 841 GTCAGTGGTGTGTTGTTATTAACAATGC-----GGTGCCCAAGTTTGGTAAACCTT 894  
DB 401 -----GlyIleIleTyrPheSerMetPhePheAsnAlaIleTrpPheGlyLeuIleValPhe 418

QY 895 GGAACACTTGGTACATACACAGATTCACAGTTCAGCTCACCGGTGGAAACTAGATT 954  
DB 419 LeuCysMetSerLeuTyrLeuThrLeuThrIleValValThrGluTrpArgThrLysPhe 438  
QY 955 AGAATGAATAAGAACAAAGCAGATTAATGATGCAGTATAGCTGTATAGACTACTGCTG 1014  
DB 439 ArgArgAlaMetAsnThrGlnGluAsnAlaThrArgAlaArgAlaValAspSerLeuLeu 458  
QY 1015 AATTATGAACCTGAGATATTTAATATGAAGATATGAAACAGACAGATATGATGA 1074  
DB 459 AsnPheGluTrpValLysTyrTyrAsnAlaGluSerTyrGluValGluArgTyrArgGlu 478  
QY 1075 TTTTGAAGAGTATGAGACTGCTTCATGAAAAGTACCTTACTGCTGCTATGCTGAAC 1134  
DB 479 AlaIleIleLysTyrGlnGlyLeuGluTrpLysSerSerAlaSerLeuValLeuAsn 498  
QY 1135 TTTGTCAAAGTGCATTTTATAGTGTGCTTAAACAGTATATAGTGCTCCGACATCAG 1194  
DB 499 GlnThrGlnAsnLeuValIleGlyLeuGlyLeuLeuAlaGlySerLeuLeuCysAlaTyr 518  
QY 1195 GGAATTGTGGCAGGATACCCCTTACTGTTGAGATCTAGTAATAGTGAATGACCTCTTTT 1254  
DB 519 PheValThrGlnGlnLysLeuGlnValGlyAspTyrValLeuPheGlyThrTyrIleIle 538  
QY 1255 CAGCTTTCATTACCCCTGAACTTCTGGAACCTGATATAGACAGACTAGACAACACTC 1314  
DB 539 GlnLeuTyrMetProLeuAsnTrpPheGlyThrTyrTyrArgMetIleGlnThrAsnPhe 558  
QY 1315 ATAGATATGAACACCTGTTACTTACTCTACTCAAGGTAGACACCCAAATTAAGACAAATG 1374  
DB 559 IleAspMetGluAsnMetPheAspLeuLeuLysGluGluThrGluValLysAspLeuPro 578  
QY 1375 ATGGCATCTCCCTTCAGATCACACACAGACACTACCGGCGCTTGTATTAATGTGAT 1434  
DB 579 GlyAlaGlyProLeuAlaArgPhe-----GlnLysGlyArgIleGluPheGluAsnValHis 596  
QY 1435 TTTGATATCATTTAGGGGACAGAAAGTCCTTAGTGAATATCTTTGAAGTCCGTCAGA 1494  
DB 597 PheSerTyrAlaAspGlyArgGluThrLeuGlnAspValSerPheThrValMetProGly 616  
QY 1495 AAGAAAGTGGCATTTGAGAGGTAGTGGTCCAGAGAAAGACAAATATGAGGCTATTA 1554  
DB 617 GlnThrLeuAlaLeuValGlyProSerGlyValGlyLysSerThrIleLeuArgLeuLeu 636  
QY 1555 TTTGCTCTTATGAGCCCTCAAAAGGTAGCACTTATCTGCTGGTCAAAATATCAAGAT 1614  
DB 637 PheArgPheTyrAspIleSerSerGlyCysIleArgIleAspGlyGlnAspIleSerGln 656  
QY 1615 GTGAGCCTGGAAAGCTTTGGAGGGCAGTGGAGTGTACTCAGATGCTGTCTTTC 1674  
DB 657 ValThrGlnAlaSerLeuArgSerHisIleGlyValValProGlnAspThrValLeuPhe 676  
QY 1675 CATATATCTATTTATTCACACTCTTATATGAAAACATCACTGCTTCACTGAGAGATG 1734  
DB 677 AsnAspThrIleAlaGluAsnIleArgTyrGlyArgValThrAlaGlyAsnAspGluVal 696  
QY 1735 TATGAGTGGCAAAATTAAGCTGAGCTTCATGATGCAATTTCTGGAATGCCCATGATAT 1794  
DB 697 GluAlaAlaAlaGlnAlaAlaGlyIleHisAspAlaIleMetAlaPheProGluGlyTyr 716  
QY 1795 GACACCCCAAGTAGGAGAAACGAGACTCAAGCTTTCAGAGAGAGAAAGCAAGATAGCA 1854  
DB 717 ArgThrGlnValGlyGluArgGlyLeuLysLeuSerGlyGlyGluLysGlnProValAla 736  
QY 1855 ATTGCAAGAGCACTTTTGAAGAGACCCCACTCACTCTATGATGAAGCTATCTCATCG 1914  
DB 737 IleAlaArgThrIleLeuLysValProGlyIleIleLeuLeuAspGluAlaThrSerAla 756  
QY 1915 TTAGATTGATTTACTGAAGACATATCTTGTCGCATGAAAGATGTGGTCAAAACACA 1974  
DB 757 LeuAspThrSerAsnGluArgAlaIleGlnAlaSerLeuAlaLysValCysAlaAsnArg 776

QY 1975 ACTTCTATTTTCATGCAACAGATGTCAGTGGTTGATGCAATGAAATCATTTCTC 2034  
Db 777 ThrrThrlleValAlaAlaHisArgLeuSerThrValAlaAsnAlaAspGlnlleuVal 796  
QY 2035 TTGGATGACGAGTAAAGTCCGACGAGTGTACCCACGATGGTTGGTCTTACCTCAT 2094  
Db 797 IlelYasArgGlyCysIleValAlaGlnArgGlyArgHisGlnAlaLeuLeuSerArg--Gly 815  
QY 2095 AGTATCTATTCAGAAATGTGCATATACAGAGACGCGTGTGACAAACATGATTAACCCC 2154  
Db 816 GlyValTyrAlaAspMetTrpGlnleuGlnGln-----GlnGlyGlnGlnGlnThr 832  
QY 2155 AATGGGAGCAAGAAAGAAATATATATCCAA 2187  
Db 833 SerGlnAspThrLysProGlnThrMetGlnArg 843  
RESULT 11  
ID AAY08961 standard; Protein; 475 AA.  
XX AAY08961;  
XX 19-AUG-1999 (first entry)  
XX A. gossypii ORF 1 encoded protein.  
XX Purine biosynthesis; phosphoribosylpyrophosphate synthetase; KPR1; KPR2;  
KM glutamine-phosphoribosylpyrophosphate amidotransferase; ADA4; GUA1;  
KM IMP-dehydrogenase; GUA1; GMP-synthetase; GUA2; PRRP synthetase; mouth;  
KM riboflavin; vitamin B2; treatment; inflammation; throat; mucosal layer;  
KM skin disorder.  
XX Aahbya gossypii.  
OS EP927761-A2.  
PN 07-JUL-1999.  
XX 08-DEC-1998; 98EP-0123331.  
XX 23-DEC-1997; 97DE-1057755.  
XX (BAD1 ) BASF AG.  
XX Hoeffken HW, Jimenez A, Pompejus M, Revuelta Doval JL;  
PI Santos Garcia MA, Seubberger H;  
XX MPI, 1999-373465/32.  
DR N-PSDB; AAX78080.  
XX Aahbya gossypii purine biosynthesis enzymes and related genes for  
PT production of riboflavin  
XX Example 3; Page 22-24; 48pp; German.  
XX This invention describes novel enzymes involved in purine biosynthesis  
CC in Aahbya gossypii. The enzymes described in the invention include  
CC Phosphoribosylpyrophosphate (PPRP) synthetases KPR1 and KPR2  
CC glutamine-phosphoribosylpyrophosphate (PPRP) amidotransferase (ADA4)  
CC IMP-dehydrogenase (GUA1) and a GMP-synthetase (GUA2). The nucleic acid  
CC sequences encoding the enzymes involved in purine biosynthesis are used  
CC to genetically modify microorganisms, for the production of riboflavin.  
CC Riboflavin, also known as vitamin B2, is essential in humans and animals  
CC and is used to treat inflammation of the mouth and throat mucosal layer  
CC as well as inflammation in skin disorders. This sequence represents the  
CC protein product of an open reading frame which flanks the ADA4 gene which  
CC encodes a glutamine PRRP amidotransferase.  
XX  
SQ Sequence 475 AA;  
Alignment Scores: 8.44e-104 Length: 475  
Pred. No.: 1201.50 Matches: 234  
Score:

Percent Similarity: 71.74% Conservative: 96  
Best Local Similarity: 50.87% Mismatches: 127  
Query Match: 28.24% Indels: 3  
DB: 20 Gaps: 2  
AF133659 (1-2345) x AAY08961 (1-475)  
QY 757 ATTGACAGAGAAACAAGGGGTATCAGTTTGTCTGAGCTTTGGTATTAATCTTCTT 816  
Db 1 MetAspArgGlyCysLysGlyIleSerTyrValLeuSerIleMetValPheHisIleIle 20  
QY 817 CCATCATGTTTGAAGTATGATCTTGTGATGTTGTTTGTATTAACAATGCGGTGCCAG 876  
Db 21 ProIleThrPheGlnIleSerMetValCysGlyIleLeuThrTyrGlnPheGlyAlaSer 40  
QY 877 TTTGGCTTTGGTAAACCTTGAGACCTTGATACATACACACATCTTCACATGTCAGTCA 936  
Db 41 PheAlaAlaIleThrPheSerThrMetLeuLeuTyrSerIlePheThrPheArgThrThr 60  
QY 937 CGGTGAGAACTAGATTATTAAGTAATGACAAAGACAGATTAATGACAGTAAATGCT 996  
Db 61 AlaTrpArgThrArgPheArgAspAlaAsnLysAlaAspAsnLysAlaAlaSerVal 80  
QY 997 GCTATGACTGCTGCTGAAATTAAGAACTGTGAAGTATTTAATAAGAAATATGAA 1056  
Db 81 AlaLeuAspSerIleuIleAsnPheGlnAlaValLysTyrPheAsnGlnLysTyrLeu 100  
QY 1057 GCACAGAGATATGATGATTTTGAAGACCTATGACAGCTGCTTCATGAAAGACCTCT 1116  
Db 101 AlaAspLysTyrHisThrSerLeuMetLysTyrArgAspSerGlnIleLysValSerGln 120  
QY 1117 ACTGTGCTATGCTGAACCTTGTCAAGTGCTATTTTTCAGTGTGCTTAAACAGCTATA 1176  
Db 121 SerLeuAlaPheLeuAsnThrGlyGlnAsnLeuIlePheThrAlaLeuThrAlaMet 140  
QY 1177 ATGGTCTGCCAGTACAGGAATTTGGCAGGTACCTTACTGTGGAGATCTTAATGAT 1236  
Db 141 MetTyrMetAlaCysAsnGlyValMetGlnGlySerLeuThrValGlyAspLeuValLeu 160  
QY 1237 GTGAATGAGTCTGTTTTCAGCTTTTCACTTACCCCTGAACCTTGTGGAACTGATATGA 1296  
Db 161 IleAsnGlnLeuValPheGlnLeuSerValProLeuAsnPheLeuGlnGlySerValTyrArg 180  
QY 1297 GAGACTAGACACAGCTCATAGATATGAACACTGTGTACTTACTCAAGGTAGACACC 1356  
Db 181 AspLeuLysGlnSerIleuIleAspMetGlnSerLeuPheLysLeuGlnLysAsnGlnVal 200  
QY 1357 CAATTAAGACAAAGTATGATGTCATCTCCCTTCAGATCAACACAGACAGTACCGTG 1416  
Db 201 ThrIleLysAsnSerProAsnAlaGlnAsnLeuProIle---HisLysProLeuAspIle 219  
QY 1417 GCCTTTGATTAATGTGATTTTGAATATCATTGAGGGCCAGAAAGTCTTATGGAATATCC 1476  
Db 220 ArgPheGlnAsnValThrPheGlyTyrAspProGlnArgGlnLeuAsnAsnValSer 239  
QY 1477 TTGAGTCCCTGACAGAAAGAAAGTGCATGTATGAGAGTGTGGGTACAGGAAAGC 1536  
Db 240 PheThrIleProAlaIleTyrMetLysThrAlaIleValGlyProSerGlySerGlyLysSer 259  
QY 1537 ACAATGATGAGCTTATTTATTTGCTTTATGAGCCTCAAAAGGTAGCATTTATCTTGCT 1596  
Db 260 ThrIleLeuLysLeuValPheArgPheTyrGlnProGlnGlnGlyArgIleLeuValGly 279  
QY 1597 GGTCAAAATATACAAATGTGAGCTGGAAGACCTTCGAGAGGCACTGGAAGTATGACT 1656  
Db 280 GlyThrAspIleArgAspLeuSerLeuLeuSerLeuArgLysAlaIleGlyValPro 299  
QY 1657 CAGATGCTGCTCTTCAATATAATCTATTTATTAACAACCTTTATGGAAGAAATCATGAT 1716  
Db 300 GlnAspThrProLeuPheAsnAspThrIleTrpGlnAsnValLysPheGlyAsnIleSer 319  
QY 1717 GCTTCACTGAGAGATGATGCAAGTGGCAAAATTAAGCTGAGACTTCATGATGCAATTTCTT 1776

|   |   |               |  |  |     |
|---|---|---------------|--|--|-----|
| D | b |               | 320  | SerSerIAspArgGluLeuLeuAlaValIleGluLysPheAlaGlnLeuThrLysLeuLeuGln | 339 |
| O | y | 1777          | CGAATGCCACATGGATATGTACAGCCCAAGTAAAGGZGAACAAGACTCAGACTTCACAGAGA         | 1836   |     |
|   |   |               | ::::                            ::::                                   |  |     |
| D | b | 340           | AsnLeuProLysGlyIleAspSerThrValIleGlyGlnArgGlyLeuMetIleSerGlyGly        | 359  |     |
| O | y | 1837          | GAAAGCAAGAGTAGGCATTGTCGAAGAGCATTTTGAAAGGACCCCCSGATATCTGTAT             | 1896   |     |
|   |   |               |  |  |     |
| D | b | 360           | GluLysGlnTrgLeuAlaIleAlaValLeuLysLysVasPalaProLeuMetPhePhe             | 379  |     |
| O | y | 1897          | GATGACGACTTCATCATGCTTAGATTCGATTACTGAAGACATCTTCGTGGCCATAAG              | 1956   |     |
|   |   |               |  |  |     |
| D | b | 380           | AspGlnAlaThrSerAlaLeuSerPheTrnHsStnGlnGlnAlaLeuLeuHisThrIleGln         | 399  |     |
| O | y | 1957          | GAT-----GTGTCAAACACAGAACTTCTATTTTCATTCGACACAGAKTTGTCAACKGT             | 2010   |     |
| D | b | 400           | GlnAsnPheSerSerAsnSerLysThrSerValTyValAlaHisArgLeuAlaGlyThrIle         | 419  |     |
| O | y | 2011          | GTTCATGCAGATGAATCATCTGTTCGATPCAGGGTAAGGTAGCCGAACGGTAAACCAC             | 2070   |     |
|   |   |               |  |  |     |
| D | b | 420           | AlaAspAlaAspLysIleIleValLeuGlnGlnGlySerValAlaGlyGlnGlyThrHis           | 439  |     |
| O | y | 2071          | CATGTTGTCGTGGTAAACCTCATATGATCTATTTCAGAAATGTGGCATAPACACAGACG            | 2130   |     |
| D | b | 440           | SerSerLeuLeuAlaSerGlnGlySerLeuTyArgGlyLeuTrpAspIleGlnGlnAsn            | 459  |     |
|   |   |               |  |  |     |
| R | E | S             | RESULT 12  |  |     |
| I | D | AAY54454      | AAY54454 standard; Protein; 574 AA.                                    |  |     |
| X | X | AA54454;      |  |  |     |
| X | X |               |  |  |     |
| D | T | 25-APR-2000   | (first entry)  |  |     |
| X | X |               |  |  |     |
| D | E |               | Amino acid sequence of a human transport protein homologue HTPH-1.     |  |     |
| X | X |               |  |  |     |
| K | M |               | Human; transport protein homologue; HTPH-1; ABC transporter; cancer;   |  |     |
| K | M |               | reproductive disorder; adenocarcinoma; leukemia; lymphoma; melanoma;   |  |     |
| K | M |               | myeloma; sarcoma; teratocarcinoma; prolactin production; infertility;  |  |     |
| K | M |               | ovarian hyperstimulation syndrome; uterine fibroid; Peyronie's disease |  |     |
| K | M |               | gynecomatia; copper metabolism disorder; Menke's disease;              |  |     |
| X | X |               | Wilson's disease; Ehlers-Danlos syndrome type IX.                      |  |     |
| O | S | Homo sapiens. |  |  |     |
| F | H | XX            |  |  |     |
| F | H | XX            |  |  |     |
| F | T | Key           | Location/Qualifiers  |  |     |
| F | T | Modified-site | 26   |  |     |
| F | T |               | /note= "potential phosphorylation site"                                |  |     |
| F | T | Modified-site | 73   |  |     |
| F | T |               | /note= "potential phosphorylation site"                                |  |     |
| F | T | Modified-site | 74   |  |     |
| F | T |               | /note= "potential phosphorylation site"                                |  |     |
| F | T | Modified-site | 89   |  |     |
| F | T |               | /note= "potential phosphorylation site"                                |  |     |
| F | T | Modified-site | 126  |  |     |
| F | T |               | /note= "potential phosphorylation site"                                |  |     |
| F | T | Modified-site | 179  |  |     |
| F | T |               | /note= "potential glycosylation site"                                  |  |     |
| F | T | Modified-site | 194  |  |     |
| F | T |               | /note= "potential phosphorylation site"                                |  |     |
| F | T | Modified-site | 230  |  |     |
| F | T |               | /note= "potential glycosylation site"                                  |  |     |
| F | T | Modified-site | 330  |  |     |
| F | T |               | /note= "potential phosphorylation site"                                |  |     |
| F | T | Modified-site | 337  |  |     |
| F | T |               | /note= "potential phosphorylation site"                                |  |     |
| F | T | Region        | 355..362   |  |     |
| F | T |               | /note= "P loop"  |  |     |
| F | T | Modified-site | 393  |  |     |
| F | T |               | /note= "potential phosphorylation site"                                |  |     |
| F | T | Modified-site | 409  |  |     |
| F | T |               | /note= "potential glycosylation site"                                  |  |     |
| F | T | Modified-site | 411  |  |     |

```

FT      Region                                /note= "potential phosphorylation site"  

FT                                          459..473  

FT                                          /note= "ABC transporter signature sequence"  

FT Modified-site                            460  

FT                                          /note= "potential phosphorylation site"  

FT Modified-site                            487  

FT                                          /note= "potential phosphorylation site"  

FT Modified-site                            491  

FT                                          /note= "potential phosphorylation site"  

FT Modified-site                            507  

FT                                          /note= "potential glycosylation site"  

FT Modified-site                            563  

FT                                          /note= "potential phosphorylation site"  

FN WQI20000301S-A2.  

PD 20-JAN-2000.  

XX XX  

PF 08-JUL-1999;    99MO-US15482.  

XX XX  

PR 10-JUN-1998;    98US-0113427.  

XX XX  

PA (INCY-) INCYTE PHARM INC.  

XX XX  

PI Hillman JL, Yie H, Reddy RM, Gorgone GA, Corley NC, Azimzai Y;  

PI Patterson C, Baughn MR;  

DR WPJ; 2000-160925/14.  

XX N-PSDB; AAZ45823.  

PT New human transport protein homologs, polypeptides and nucleic acids  

PT useful for diagnosing, treating or preventing cancer, reproductive  

PT disorders and copper metabolism disorders -  

XX XX  

PS Claim 1; Page 68-69; 75pp; English.
```

The present sequence represents a human transport protein homologue,  
designated HTPH-1. HTPH-1 is a member of the ABC transporter family.  
The antagonists of the protein are useful for treating or preventing  
cancer and reproductive disorders such as adenocarcinoma, leukaemia,  
lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, etc., and  
disorders of prolactin production, infertility, ovarian hyperstimulation  
syndrome, uterine fibroids, Peyronie's disease, gynaecomastia, etc.. The  
polynucleotide, antibodies, agonists, complementary sequences or  
vectors are also useful for treating or preventing the diseases. The  
polypeptide is also useful for treating or preventing a copper  
metabolism disorders such as Menke's disease, Wilson's disease and  
Ehler-Danlos syndrome type IX. The antibodies are also useful for  
diagnosing disorders characterized by expression of the human transport  
protein homologue.

```
SQ Sequence       574 AA;
```

Alignment Scores:

|                               |                   |                 |
|-------------------------------|-------------------|-----------------|
| Pred. No.: .....              | Score: 5.03e-102  | Length: 574     |
| Percent Similarity: 62.12%    | Matches: 1183/00  | Mismatches: 250 |
| Best Local Similarity: 42.66% | Conservative: 114 | Indels: 194     |
| Query Match: 27.81%           | Gaps: 21          | Indels: 28      |

AF133659 (1-2345) x AAY54454 (1-574)

```
OY 430 CTGGAGATTTTGGGTGTCGAAGCCCATGATTGTGCCCTTGATTTAAATAAT 489
   |||||          :|||||:|||||:|||||          :::::
Db  1 MetgylleuwetglyValargalhalenavalphevalProlhehyrnyrg--- 19
OY 430 GCTGTAGACAGCCTCAACGCAGATGTGGGAAAACATCGTAACCTG-----AGTAGACA 543
   ||||           |:|||
Db  20 -----AsnilleValasluenduehrGlulyala 29
OY 544 CCA--AATACAGTTGCACCACATGGCAAACGCA-----GTTGTGATTGGC 585
   |||         |||:|||||        |||:|||||
```



Db 30 ProTrpAsnSerLeuAlaTrpThrValThiSerTyrValPheLeuLysPheLeuGlnGly 49  
QY 586 TATGATGATATCAGAGCTGAGCTGCTTTTATTAAGAAAGTTCGAATGACATATTTGGC 645  
Db 50 GlyGlyThrGlySerThrGly-----PheValSerAsnLeuAlaGlnThrPheLeuTrpIle 67  
QY 646 AAGGTAGCCCAAGATTCAATCCGAAGATAGCCAAAAATGCTTTCTTCATCTTCACAC 705  
Db 68 ArgValGlnGlnPheThrSerArgArgValGlnLeuLeuIlePheSerHisLeuHisGln 87  
QY 706 CTGATGCTGGGTTTTCACCTGAGCAGACAGGGAGCTTATCTATGAGGCTATTCAGACA 765  
Db 88 LeuSerLeuAlaGlnTrpHisLeuGlyArgArgThrGlyGlnValLeuArgIleAlaAspArg 107  
QY 766 GAACAAGAGGGATACAGTTTGTCTCAGTCTTGATTTAACTTCTTCCCATCATG 825  
Db 108 GlyThrSerSerValThrGlyLeuLeuSerTyrLeuValPheAsnValIleProThrLeu 127  
QY 826 TTGGAAGTATGCTTTCAGTGTGCTTTTGTATTAACAATCC-----GGTCCCACTT 879  
Db 128 AlaAspIleIleIle-----GlyIleIleTyrPheSerMetPhePheAsnAlaTrpPhe 145  
QY 880 GCTTTGGTAACCTTGGAAACCTGTGATACACAGATTCAGATTCAGTTCAGTACACGG 939  
Db 146 GlyLeuIleValPheLeuCysMetSerLeuTyrLeuThrIleValIleValThrGln 165  
QY 940 TGGAGAACTAGATTAGAAATGAAATGAAACAACAGATATGATGACGATATGCTGT 999  
Db 166 TrpArgThrLysPheArgArgAlaMetCysThrGlnGlnAsnAlaThrArgAlaArgAla 185  
QY 1000 ATGACTCTACTGCTGAATTATGAAACCTGTGAAGTATTTAAATGAAAGATATGAA 1059  
Db 186 ValAspSerLeuLeuAsnPheGlnThrValLysTyrTrpAsnAlaGlnSerTyrGlnVal 205  
QY 1060 CAGAGATATGATGATTTTGAAGACGATAGACGCTTCAATTGAAAGTACTTACT 1119  
Db 206 GlnArgTyrArgGlnAlaIleIleLysTyrGlnGlnLeuGlnLysSerSerAlaSer 225  
QY 1120 CTGGCTATGCTGAACCTTGTGCAAGTGTCTTATTTCACTGCTGCTTAAACGATATATG 1179  
Db 226 LeuValLeuLeuAsnGlnThrGlnAsnLeuValIleGlyLeuGlyLeuLeuAlaGlySer 245  
QY 1180 GTGCTGCCCACTGAGGAATTGTGGACAGTACCTTACTGTGGAGATCTAGTAATGTG 1239  
Db 246 LeuLeuCysAlaTyrPheValThrGlnGlnLysLeuGlnValGlyAspTyrValLeuPhe 265  
QY 1240 AATGCACTGCTTTTTCAGCTTTTCAATACCCCTGAACCTTCTGGAACTGATATAGAG 1299  
Db 266 GlyThrTyrIleIleGlnLeuTyrMetProLeuAsnTrpPheGlyThrTyrTyrArgMet 285  
QY 1300 ACTGACACAGCACTCATGATATGACACCTGTGTTTCTACTACAGGATAGACCCAA 1359  
Db 286 IleGlnThrAsnPheIleAspMetGlnAsnMetPheAspLeuLysGlnGlnThrGln 305  
QY 1360 ATTAAGACAAAGTATGCGATCTCCCTTCAGATCAACACACAGCAGTACCGTGCC 1419  
Db 306 ValLysAspLeuProGlyAlaGlyProLeuArgPhe-----GlnLysGlyArgIleGln 323  
QY 1420 TTTCATATATGCTTTTGAATACATTGAGGGCCAGAAAGTCTTACTGCAATATCTTT 1479  
Db 324 PheGlnAsnValHisPheSerTyrAlaAspGlyArgGlnThrLeuGlnAspValSerPhe 343  
QY 1480 GAAGTCCCTGAGGAAGAAAGATGCGCATTTAGAGTGTAGGTGAGGAGAAAGACA 1539  
Db 344 ThrValMetProGlyGlnThrLeuAlaLeuValGlyProSerGlyAlaGlyLysSerThr 363  
QY 1540 ATAGTGGCTATTATTTCCTTCTATAGCGCTCAAAAGGATAGCATTTATCTGTGCT 1599  
Db 364 IleLeuArgLeuLeuPheArgPheTyrAspIleSerSerGlyCysIleArgIleAspGly 383  
QY 1600 CAAATATACAAAGTGTGAGCTTGAAAGCTTTGAGAGGCAGTGGAGATGCTTCTCAG 1659  
Db 384 GlnAspIleSerGlnValThrGlnAlaSerLeuArgSerHisIleGlyValValProGln 403

QY 1660 GATGCTGTCTCTTCCATTAATACATTTATTAACAACCTTATATGGAACATCAGTCT 1719  
Db 404 AspThrValLeuPheAsnAspThrIleAlaAspAsnIleArgTyrGlyArgValThrAla 423  
QY 1720 TCACCTGAGGAAGTATGATGAGTGGCAAAATTTGCTGAGCTTCAATGATATCTTCCA 1779  
Db 424 GlyAsnAspGlnValGlnAlaAlaGlnAlaGlnAlaGlyIleHisAspAlaIleMetAla 443  
QY 1780 ATGCACATGATATATGACACCAAGTAGGGGAGAGAGACTCAAGCTTCAGAGAGAA 1839  
Db 444 PheProGlnGlyTyrArgThrGlnValGlyIuArgGlyLeuLysLeuSerGlyGlyGln 463  
QY 1840 AAGCAAAAGATAGCAATTTGCAAGAGCCATTTTGAAGAGCCCCCAGTCACTATGAT 1899  
Db 464 LysGlnArgValAlaIleAlaIleArgThrIleLeuLysAlaProGlyIleIleLeuLeuAsp 483  
QY 1900 GAAGCTATCTTCACTGCTTACATTCGATTAAGAGACATATTTGGTGGCAGTAAAGAT 1959  
Db 484 GlnAlaThrSerAlaLeuAspThrSerAsnGlnArgAlaIleGlnAlaSerLeuAlaLys 503  
QY 1960 GTGGTAAACAGAGAACTTCTATTTTCATTGACACAGATTCGACAGTGTGATGCA 2019  
Db 504 ValCysHisAsnArgThrThrIleValAlaHisArgLeuSerThrValValAsnAla 523  
QY 2020 GATGAATCATTTGCTTGGATCAGGTAAAGGTAGCCGAACGTGTGATCCCATGCTTTG 2079  
Db 524 AspGlnIleLeuValIleLysAspGlyCysIleValGlnArgGlyArgHisGlnAlaLeu 543  
QY 2080 CTTCCTAACCTCTATATGATCTTATTCAGAAATGTGGCATACACAGACCCGTGTGAG 2139  
Db 544 LeuSerArg---GlyGlyValTyrAlaAspMetTrpGlnLeuGlnGlnGlnGln 562  
QY 2140 AACCATGATTAACCCCAA 2157  
Db 563 ThrSerGlnAspThrLys 568  
RESULT 13  
ID ABB71210 standard; Protein; 866 AA.  
XX ABB71210;  
AC ABB71210;  
XX 26-MAR-2002 (first entry)  
DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 40422.  
XX Drosophila, developmental biology; cell signalling; insecticide;  
XX Drosophila, developmental biology; cell signalling; insecticide;  
XX Drosophila melanogaster.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX 27-SEP-2001.  
PD 27-SEP-2001.  
PF 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2001; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
XX PA Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-65660/75.  
XX DR N-PSDB; ABL15313.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX PS Disclosure; SEQ ID NO 40422; 21bp + Sequence Listing; English.



Db 813 LeuLysAsnLeuAspAlaGluGlnSerGlyGlySerAspAsnGlyAspAlaSerAlaGlu 832  
QY 2176 AATATATCCAAAAGAGAGAAAGAAAG 2202  
Db 833 -----SerglySerglyLysArgArg 839  
RESULT 14  
AAM39313  
ID AAM39313 standard; Protein; 571 AA.  
XX  
AC AAM39313;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 2458.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW Leukemia.  
XX  
OS Homo sapiens.  
XX  
PN MO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000MO-US34263.  
XX  
PR 21-JAN-2000; 2000US-048725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
XX  
DR N-PSDB; AAI58469.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 4; SEQ ID NO 2458; 10078bp; English.  
XX  
XX  
CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cyostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localized neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 571 AA;

Alignment Scores:  
Pred. No.: 3,51e-101 Length: 571  
Score: 1174.00 Matches: 248  
Percent Similarity: 62.26% Conservative: 115  
Best Local Similarity: 42.54% Mismatches: 192  
Query Match: 27.60% Indels: 28  
DB: 22 Gaps: 9  
AF133659 (1-2345) x AAM39313 (1-571)  
QY 439 TTGGGTGGTGCAGAAAGCCATGAAATATTGTCCTTCATGTTAAATAGCTGATAGC 498  
Db 1 MetGlyLeuGluArgAlaLeuAsnValLeuAlaProIlePheTyrArg-----16  
QY 499 AGCCTCAACACAGATGTCGGAAACATGCTGAACTG-----AGTAGACACCA---AAT 549  
Db 17 -----AsnIleValAsnLeuLeuThrGluAsnAlaProTyrPheN 29  
QY 550 ACAGTTCAGAACCATGACAGACAGCA-----GTTCTGATTGGCTATGGTGA 594  
Db 30 SerLeuAlaIleThrValThrSerTyrValPheLeuLysPheLeuGlnGlyGlyThr 49  
QY 595 TCAGAGAGCTGGAGCTCTTTTAAAGAGATTCCGAATGCAAGTATTTGGCAAGTACC 654  
Db 50 GlySerThrGly-----PheValSerAsnLeuArgThrPheLeuThrPleArgValGln 67  
QY 655 CAGAAATTCAGATCCAGAGAAATAGCCAAATAATGCTTCATTCACAACTGGATTCG 714  
Db 68 GlnPheThrSerArgArgValGluLeuLeuIlePheSerHisLeuHisGluLeuSerLeu 87  
QY 715 GGTGTTTCACCTGAGACAGACAGACGGAGCTTATATCAAGCTATTGACAGAGACAG 774  
Db 88 ArgTrpHisLeuGlnYargTrgTrgGluValLeuArgIleAlaLeuArgGlyThrSer 107  
QY 775 GGTATGAGTTTGTCTGAGTGTCTTGGATTTAACTTCTTCCATCATGTTGTAAGTG 834  
Db 108 SerValThrGlyLeuLeuSerTyrLeuValPheAsnValIleProThrLeuAlaAspIle 127  
QY 835 ATGCTGTGTCAGTGGTGTGTTGTTGTTATCAATGC-----GGTCCCAAGTTTGGTGA 888  
Db 128 IleIle-----GlyIleIleTyrPheSerMetPhePheAsnAlaTrpPheGlyLeuIle 145  
QY 889 ACCCTTGAACACTTGTTGATACATACAGCATTCACAGTTGCAGTGCACAGCTGAGAACT 948  
Db 146 ValPheLeuCysMetSerLeuTyrLeuThrLeuThrIleValValThrGluTrpArgThr 165  
QY 949 AGATTGATTAAGTAATGAATGAACAAAGCAGTAATGATGACAGTAACTGCTATAGACTA 1008  
Db 166 LysPheArgArgAlaMetAsnThrGlnGluAsnAlaThrArgAlaArgAlaValAspSer 185  
QY 1009 CTGCTGAATTAAGAACTGTAAGTATTTTAATTAATGAAGATATGACAGACAGAAAT 1068  
Db 186 LeuLeuAsnPheGluThrValIleTyrTyrIleAsnAlaGluSerTrgIleValGluArgTyr 205  
QY 1069 GATGATTTTGAAGACGTATGAGACTGCTTATGAAAGTAAAGTAACTGCTGAGTAAAG 1128  
Db 206 ArgGluAlaIleIleLeuTyrGlnGlyLeuGluTrpLysSerLeuAsnValLeu 225  
QY 1129 CTGAACCTTGTGTAAGAGCTATTTTCACTGCTGTTTAAAGCTATTAAGTGTCTGCC 1188  
Db 226 LeuAsnGlnThrGlnAsnLeuValIleGlyLeuGlyLeuLeuAlaGlySerLeuLeuCys 245  
QY 1189 AGTCAGGAAATGTGCGACAGTACCTTACTGTTGAGATCTGATATGATGATGACTG 1248  
Db 246 AlaTyrPheValThrGlnGlnLysLeuGlnValGlyAspTyrValLeuLeuGlyThrTyr 265  
QY 1249 CTTTTCACCTTCAATTAACCTGAACTTTCGGAACCTGATATGAGAGACATGACAA 1308  
Db 266 IleIleGlnLeuTyrMetProLeuAsnTrpPheGlyThrTyrTrpArgMetIleGlnThr 285  
QY 1309 GCACCTATATGATATGAAACACCTTGTACTTACTCAAGTATGACCAACCAATTAAGAC 1368

```

Db 286 AsnPhel1eAspMetGluAsnMetPheAspLeuLeuLysGluGluThrGluValLysAsp 305
QY 1369 AAGAGTAGGAGCTCCCTTCAGATCACACAGACAGCTACCGCTTGATAT 1428
Db 306 LeuPProGlyAlaGlyProLeuLysPhe-----GlnLysGlyAlaGlyLeuPheGluAsn 323
QY 1429 GTGCAATTTGATACATTGAGGCGCAAGAAAGCTTAACTGAAATATCTTTGAACTCCCT 1488
Db 324 ValHisPheSerTyrAlaAspGlyArgGluThrLeuGlnAspValSerPheThrValMet 343
QY 1489 GCAGAGAAAGAAAGTGGCCATTGTAGAGTAGTGGCTCAGGGAAAGACAACTAGTGAGC 1548
Db 344 ProGlyGlnThrLeuAlaLeuValGlyProSerGlyAlaGlyLysSerThrLeuLysArg 363
QY 1549 CTATATTTGCTTCTATAGAGCTCAAAAGGGTAGCAATTTATCTTCGCTGGTCAAAATATA 1608
Db 364 LeuLeuPheArgPheTyrAspLysSerGlyCysIleArgIleAspGlyAsnAsp 383
QY 1609 CAAGATGTGAGCCTTGAAAGCCTTCGAGGGGAGGAGTGGAGTGAAGTCTCAGAGATCTGTC 1668
Db 384 SerGlnValThrGlnAlaSerLeuLysSerHisIleGlyValValProGlnAspThrVal 403
QY 1669 CTCCTTCATATATCTATTTATTTACAACTCTTATATGAAACATGATGCTTCACCTGAG 1728
Db 404 LeuPheAsnAspThrIleAlaAspAsnIleArgTyrGlyArgValThrAlaGlyAsnAsp 423
QY 1729 GAAGTGTATGAGTGGCAAAATTTAGTCACTTCAATGCAATTTCTGCAATGCAACT 1788
Db 424 GluValGluAlaAlaAlaGlnAlaAlaGlyIleHisAspAlaIleMetAlaPheProGlu 443
QY 1789 GGATATGACACCCCAAGTAGGAGGAGACGAGACTCAAGCTTTGAGAGAGAAAGCAAGA 1848
Db 444 GlyTyrArgThrGlnValGlyGluArgGlyLeuLysSerGlyGlyGluLysGlnArg 463
QY 1849 GTAGCAATTTGCAAGACCACTTTTGAAGACCCCAAGTCAATCTTATGATGAGCTACT 1908
Db 464 ValAlaIleAlaArgThrIleLeuLysAlaProGlyIleIleLeuLeuAspGluAlaThr 483
QY 1909 TCATGTTAGTATGATTAAGTATTAAGACACTATTTCTGGCCATGAGAGTGTGTCAA 1968
Db 484 SerAlaLeuAspThrSerAsnGlyArgAlaIleGlnAlaSerLeuAlaLysValCysAla 503
QY 1969 CACAGAACTCTTATTTTTCATTCACACAGATTTGCAACAGTGGTGGTACAGATGAAATC 2028
Db 504 AsnArgThrThrIleValValAlaHisArgLeuSerThrValValAlaHisLysAspGlnIle 523
QY 2029 ATTGCTTGATCAGGAGTAAAGTACCCCAAGCTGTGATCCCACTGATGTTGCTTCTAAC 2088
Db 524 LeuValIleLysAspGlyCysIleValGluArgGlyArgHisGluAlaLeuLeuSerArg 543
QY 2089 CCTCATAGTATCTATTCGAATGTGGCATACACAGAGACCGGTGAGAAACATGAT 2148
Db 544 ---GlyGlyValTyrAlaAspMetTyrGlnLeuGlnGlnGlyGlnGluThrSerGlu 562
QY 2149 AACCCCAAA 2157
Db 563 AspThrLys 565

```

```

KW 1eukemia.
XX
OS Homo sapiens.
XX
PN W0200153312-A1.
XX
PD 26-JUL-2001.
XX
PE 26-DEC-2000; 2000WO-US34263.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB: AA160255.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 6030; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA138642-AA142213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC activation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SO Sequence 591 AA;

```

```

Alignment Scores:
Pred. No.: 7.14e-99 Length: 591
Score: 1149.50 Matches: 247
Percent Similarity: 61.14% Conservative: 118
Best Local Similarity: 41.37% Mismatches: 203
Query Match: 27.02% Indels: 29
DB: Gaps: 10

```

AF133659 (1-2345) x AAM41099 (1-591)

```

QY 400 CCAGATCTAGAGAGCTAGAGTTCGCTGGATTTTGGTGGTGAAGAGCCATG 459
Db 7 ProAlaLeuGlnLeuGluValLeuIleCysLeuGlyLeuMetGlyLeuGluArgAlaLeu 26
QY 460 AATATGCTGCTCCCTCATGTTAAATATGCTGTAGACGCTCAACCAAGATGCCGA 519
Db 27 AsnValLeuAlaProIlePheTyrArg----- 35
QY 520 AACATGCTGAACCTG-----AGTATGACCA---AATACAGTTGGACACCATGCAACA 570
Db 36 AsnIleValAsnLeuLeuThrGluAsnAlaProTyrAsnSerLeuAlaIleThrValThr 55

```

QY 571 GCA-----GTTCTGATGGCTATGCTGATACAGAGCTGAGCTGCTTTT 615  
 Db 56 SerTyrValPheLeuLysPheLeuGlnGlyGlyThrGlySerThrGly-----Phe 73  
 QY 616 TTTAACGAGATTCCGAATGAGTATTTGGCAAGTAGCCCGAATTCAATCCGAAGATA 675  
 Db 74 ValSerLeuArgThrPheLeuTrpIleArgValGlnGlnPheThrSerArgVal 93  
 QY 676 GCCAAATGCTTTCATCTTCACAACTGGATCGGGTTTCACTGAGAGAGCAG 735  
 Db 94 GluLeuLeuIlePheSerHisLeuHisGluLeuSerLeuArgTrpHisLeuGlyArgArg 113  
 QY 736 ACGGAGCTTTATCTAAAGCTATTGACAGAGAAACAAGGGATCAAGTTTGTCTGAGT 795  
 Db 114 ThrGlyValLeuArgIleAlaAspArgGlyThrSerSerValThrGlyLeuLeuSer 133  
 QY 796 GCTTTGGATTATATCTTCTCCATCATGTTGAAGTGAAGTGTGCAGGTGTTTG 855  
 Db 134 TyrLeuValPheAsnValIleProThrLeuAlaAspIleIleIle-----GlyIleIle 151  
 QY 856 TATTACAAATGC-----GGTGGCCAGTTTGTGGTAAACCTTGAAACCTTGATACA 909  
 Db 152 TyrPheSerMetPhePheAsnAlaTrpPheGlyLeuIleValPheLeuCysMetSerLeu 171  
 QY 910 TACACAGCATTCACAGTGCAGTGCACACGAGTGAACCTAGATTGAATGAATGAAC 969  
 Db 172 TyrLeuThrLeuThrIleValIleValThrGlyTrpArgThrLysPheArgArgAlaMetAsn 191  
 QY 970 AAAGCAGATTAATGATGCAAGGTAAGTCTGCTATGACTGCTGCAATTATGAACGTGTG 1029  
 Db 192 ThrGlnGlnAsnAlaThrArgAlaArgAlaValAlaSerLeuLeuAsnPheGlnThrVal 211  
 QY 1030 AAGTATTTTATTAATGAAGATATGAAGCAACAGATATGATGATTTTGAACAGTAT 1089  
 Db 212 LysTyrTyrAsnAlaGluSerTyrGluValGluArgTyrArgGluAlaIleIleLysTyr 231  
 QY 1090 GAGACTGCTTCAATGAAAGTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1149  
 Db 232 GlnGlyLeuGlnTrpLysSerSerAlaSerLeuValLeuLeuAsnGlnThrGlnAsnLeu 251  
 QY 1150 ATTTTCAGTCTGCTTAAACGCTATTAATGCTGCTGCCAGTCAAGGAAATTTGGCAGGT 1209  
 Db 252 ValIleGlyLeuGlyLeuValAlaGlySerLeuLeuCysAlaTyrPheValThrGlnGln 271  
 QY 1210 ACCCTTACTGTTGAGATCTGATATGCTGATAGTACTGCTTTTTCAGCTTTTCAATACC 1269  
 Db 272 LysLeuGlnValGlyAspTyrValLeuPheGlyThrTyrIleIleGlnLeuTyrMetPro 291  
 QY 1270 CTGAACCTTTCTGGAAGCTGTATATAGAGAGACTGACACAGCAGCATATGATGAACACC 1329  
 Db 292 LeuAsnTrpPheGlyThrTyrTyrArgMetIleGlnThrAsnPheIleAspMetGluAsn 311  
 QY 1330 TTGTTTACTCTACTCAAGGATGACACCCAAATTAAGAACAAGTGAATGCTTCCCTT 1389  
 Db 312 MetPheAspLeuLeuLysGlyThrGlnValLysAspLeuProGlyAlaGlyProPhe 331  
 QY 1390 CAGATCACACACAGACAGTACCGTGGCTTGTATATGTCATTTTGAATATATGAG 1449  
 Db 332 ArgPhe-----GlnLysGlyArgIleGlnPheGlnAsnValHisPheSerTyrAlaAsp 349  
 QY 1450 GGCAGAAAGCTCTTAAGTGAATATCTTTGAAGTCCCTGACGAGAAAGAAAGTGGCATT 1509  
 Db 350 GlyArgGlnTrpLeuGlnAspValSerPheThrValMetProGlyGlnThrLeuAlaLeu 369  
 QY 1510 GTAGAGAGTAGGGTACGAGGAGAAAGACAATAGTGAAGGCTATTATTTGCTTATGAG 1569  
 Db 370 ValGlyProSerGlyAlaGlyLysSerThrIleLeuArgLeuLeuPheArgPheTyrAsp 389  
 QY 1570 CCTCAAAAGGTACATTTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1629  
 Db 390 IleSerSerGlyCysIleArgIleAspGlyGlnAspIleSerGlnValThrGlnAlaLeu 409

QY 1630 CTTCCGAGGGCA--GTGGAGTGTGTAACCTCAGAGATGCTGTCTTCTTCAATAACTATT 1686  
 Db 410 PheArgPheSerHisTrpGlnLeuCysProLysAspThrValLeuPheAsnAspThrIle 429  
 QY 1687 TATTACACCTCTTATATGGAACATCAGCTTCCACTGAGAGAGTATGCAAGTGGCA 1746  
 Db 430 AlaAspAsnIleArgTyrGlyArgValThrAlaGlyAsnAspGlnValGluAlaAla 449  
 QY 1747 AAATTAGCTGACTCATGATGCAATTTCTGCAATCCACATGATGATGACACCAAGTA 1806  
 Db 450 GlnAlaAlaGlyIleHisAspAlaIleMetAlaPheProGlnLysTyrArgThrGlnVal 469  
 QY 1807 GGGGAACGAGACTCAAGCTTTCAGAGAGAAAGCAAGAAAGTACGATTCGAAGGCC 1866  
 Db 470 GlyGlnArgGlyLeuLysLeuSerGlyGlyGlnLysGlnArgValAlaIleAlaArgThr 489  
 QY 1867 ATTTTGAAGACCCCGACATACCTATGATGAAGCTTCAATGCTTATGATTCATT 1926  
 Db 490 IleLeuLysAlaProGlyIleIleLeuLeuAspGlnAlaThrSerAlaLeuAspThrSer 509  
 QY 1927 ACTGAAGACTATTTCTTGGTGCATGAAGATGTGGTCAACAGAACTTCTATTTTC 1986  
 Db 510 AsnGlnArgAlaIleGlnAlaSerLeuAlaLysValCysAlaAsnArgThrThrIleVal 529  
 QY 1987 ATTGCACACAGATTGTCACAGTGTGATGACAGATGAATTCATTTGATCAGGCT 2046  
 Db 530 ValAlaHisArgLeuSerThrValAlaAsnAlaAspGlnIleLeuValIleLysAspGly 549  
 QY 2047 AAGTACCCGAACGTGTGATCCCAACATGCTTGTCTGCTCAACCTCATAGTATCTATCA 2106  
 Db 550 CysIleValGlnArgGlyArgHisGlnAlaLeuLeuSerArg--GlyGlyValTyrAla 568  
 QY 2107 GAATGTGCATACACAGAGCAGCCGTTGTCAGAACCATGATATACCCANA 2157  
 Db 569 AspMetTrpGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 585

Search completed: February 15, 2003, 05:53:54  
 Job time : 91 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2003, 01:40:25 ; Search time 3255 Seconds  
(without alignments)  
11667.716 Million cell updates/sec

Title: AF133659

Perfect score: 2345  
Sequence: 1 ATGGCGCTGCTCGCATGCA.....TATTAAAAATCATACATT 2345

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1\_0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estdb:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_man:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description        |
|------------|--------|-------------|--------|-------|--------------------|
| 1          | 2277.6 | 97.1        | 2384   | 11    | AF038950 Homo sapi |
| 2          | 1061.2 | 45.3        | 1066   | 11    | U66679 Homo sapien |
| 3          | 853.8  | 36.4        | 1120   | 13    | BM479048 AGENCOURT |
| 4          | 700.4  | 29.9        | 730    | 14    | BM723815 UT-E-EOI- |
| 5          | 696.6  | 29.7        | 787    | 12    | BG722394 602693606 |
| 6          | 685.8  | 29.2        | 892    | 13    | BI463667 603207337 |

|    |       |      |     |    |          |                    |
|----|-------|------|-----|----|----------|--------------------|
| 7  | 655.2 | 27.9 | 704 | 9  | AU127740 | AU127740 AU127740  |
| 8  | 653.8 | 27.9 | 822 | 12 | BE876927 | BE876927 601484702 |
| 9  | 647.8 | 27.6 | 840 | 13 | BI549793 | BI549793 603194529 |
| 10 | 645   | 27.5 | 666 | 14 | BM975299 | BM975299 UT-CF-EC1 |
| 11 | 645   | 27.5 | 683 | 12 | BG109895 | BG109895 602279565 |
| 12 | 633.4 | 27.0 | 722 | 12 | BG722908 | BG722908 602695410 |
| 13 | 623.2 | 26.6 | 817 | 12 | BF310056 | BF310056 601892556 |
| 14 | 614   | 26.2 | 758 | 10 | BE618331 | BE618331 601462724 |
| 15 | 610.4 | 26.0 | 627 | 9  | AI769898 | AI769898 wj30C01.x |
| 16 | 604.4 | 25.8 | 638 | 14 | BM988352 | BM988352 UT-H-DT1- |
| 17 | 587.2 | 25.0 | 680 | 10 | BB204009 | BB204009 BB204009  |
| 18 | 578   | 24.6 | 602 | 14 | BQ630948 | BQ630948 111411.Y  |
| 19 | 576.6 | 24.6 | 630 | 9  | AU136079 | AU136079 AU136079  |
| 20 | 571.6 | 24.4 | 618 | 9  | AA668992 | AA668992 aa81C09.s |
| 21 | 571   | 24.3 | 572 | 12 | BG811020 | BG811020 602766741 |
| 22 | 569.8 | 24.3 | 811 | 12 | BG069454 | BG069454 H3075G12- |
| 23 | 562   | 24.0 | 564 | 10 | AM721773 | AM721773 xsl0B12.x |
| 24 | 561   | 23.9 | 589 | 10 | BE274469 | BE274469 601120316 |
| 25 | 561   | 23.9 | 611 | 9  | AA642411 | AA642411 ns27B08.s |
| 26 | 551.6 | 23.5 | 620 | 9  | AU127979 | AU127979 AU127979  |
| 27 | 548.6 | 23.4 | 595 | 10 | BE086841 | BE086841 QV1-BT067 |
| 28 | 547.2 | 23.3 | 700 | 12 | BG122119 | BG122119 602349556 |
| 29 | 546.2 | 23.3 | 579 | 10 | AA960180 | AA960180 EST372251 |
| 30 | 545   | 23.2 | 546 | 9  | AA305099 | AA305099 EST176101 |
| 31 | 542   | 23.1 | 562 | 13 | BM662767 | BM662767 UT-E-C11- |
| 32 | 541.8 | 23.1 | 559 | 9  | AA912934 | AA912934 o133b11.s |
| 33 | 540.8 | 23.1 | 609 | 9  | AA887445 | AA887445 o141a04.s |
| 34 | 530   | 22.6 | 546 | 12 | BG311594 | BG311594 na36e04.  |
| 35 | 528.6 | 22.5 | 583 | 13 | BM030723 | BM030723 159206.MA |
| 36 | 525   | 22.4 | 549 | 9  | AI016955 | AI016955 ou27F08.x |
| 37 | 520.4 | 22.2 | 546 | 10 | AM104620 | AM104620 xc84D10.x |
| 38 | 518   | 22.1 | 534 | 9  | AA923765 | AA923765 om39070.s |
| 39 | 516.2 | 22.0 | 545 | 9  | AA521270 | AA521270 aa75f07.s |
| 40 | 511.4 | 21.8 | 605 | 9  | AU150336 | AU150336 AU150336  |
| 41 | 511   | 21.8 | 513 | 9  | AI128014 | AI128014 qc42902.x |
| 42 | 507.2 | 21.6 | 559 | 10 | BB635137 | BB635137 BB635137  |
| 43 | 506   | 21.6 | 517 | 9  | AA626765 | AA626765 ab52d07.t |
| 44 | 505.4 | 21.6 | 544 | 9  | AA27515  | AA27515 z18d11.s   |
| 45 | 504.2 | 21.5 | 589 | 9  | AI678722 | AI678722 lu69c01.x |

## ALIGNMENTS

RESULT 1  
AF038950  
LOCUS AF038950  
DEFINITION Homo sapiens ATP binding cassette transporter mRNA, complete cds.  
ACCESSION AF038950  
VERSION AF038950.1 GI:3329369  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens.  
REFERENCE  
AUTHORS Mao, M., Fu, G., Wu, J.-S., Zhang, Q.-H., Zhou, J., Kan, L.-X., Huang, Q.-H., He, K.-L., Gu, B.-W., Han, Z.-G., Shen, Y., Yu, Y.-P., Xu, S.-H., Wang, Y., Chen, S.-J., and Chen, Z. Identification of genes expressed in human CD34(+) hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (14), 8175-8180 (1998)  
MEDLINE 98318631  
PUBMED 963160  
REFERENCE 2 (bases 1 to 2384)  
AUTHORS Zhang, Q.-H., Ye, M., Wu, X.-Y., Ren, S.-X., Zhao, M., Zhao, G.-J., Fu, G., Shen, Y., Pan, H.-Y., Lu, G., Zhong, M., Xu, X.-R., Han, Z.-G., Zhang, J.-W., Tao, J., Huang, Q.-H., Zhou, J., Hu, G.-X., Gu, J., Chen, S.-J., and Chen, Z. Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells  
JOURNAL Genome Res. 10 (10), 1546-1560 (2000)



MEDLINE 20499367  
PUBMED 11042152  
REFERENCE 3 (bases 1 to 2384)  
AUTHORS Zhang, Q.  
TITLE Direct Submission  
JOURNAL Submitted (16-DEC-1997) Rui-Jin Hospital, Shanghai Second Medical University, Shanghai Institute of Hematology, 197 Rui-Jin Road II, Shanghai 200025, P. R. China

FEATURES  
source 1..2384  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xq21.3"  
/cell\_type="CD34+ cell"  
/tissue\_type="cord blood"  
9..2267

CDS

/codon\_start=1  
/product="ATP binding cassette transporter"  
/protein\_id="AAC39865.1"  
/db\_xref="GI:3329370"  
/translation="MALHMSMRMAAAAFKKRHSALILIRPLVSVSGSPQMRPH  
OLGALGTRARQIPESLKSITWRLKGNSSQFLDAKALQWPLIEKRTGHHAGS  
GLHTDPKGLADVDTKIKAMLSITWPKDPDLARVPISLGFVGAKANIVPEM  
PKYAVDSLQMSGMNLSDAPNTVAATAVLIGYVSRAAEFNEVRNAPGVKA  
ONSIRIANKVFLHLNLDLGFHLSRQTALSKAIDRTGRLSEVLSLVNPPLNHV  
EMLLSVGLYKCCALLGNLGLTATFAVTRWRTRFLEIDQADNAAAI  
DSLNTVERKYFNERRVERAYDGLKTYETASIKSTTLMLNPGOSLIFSVGLTAL  
MVLASQIVACTILVVDLVMVNGHLFQSLPLNPLCTVYRTRQALIDMTLFLVLY  
DIOIKDVMASPLQTPOTATVAFDNVFEIIEGQKVLISFVEVPACKVAIVGSG  
SGKSTIVLRFYEPQKSIYLAGQNLQDVLSLRVAVNVQDAVLFNNTLYNL  
LYGNISAPBEVYAKLAGLHDAILRMPHVDIOVGRGKLSGEGRQVARIALL  
KDPVILIDEATSSLDISTEETILGAMDVVKHRSIPILRLSTVDADEIIVLDQ  
KVAREGTHGLANPHSIIYSMMHTQSRVQNHNPKEAKENISKEEKXQBEI  
VNSYKGCNCSC"

BASE COUNT 719 a 467 c 567 g 631 t  
ORIGIN

Query Match 97.1%; Score 2277.6; DB 11; Length 2384;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 2318; Conservative 0; Mismatches 24; Indels 6; Gaps 2;

QY 1 ATGGCGCTGCTGGAGATGATCTTGGCGCTGGCGCGCGCGCTGCTTTGAAAAG 60  
Db 9 ATGGCGCTGCTGGAGATGATCTTGGCGCTGGCGCGCGCGCTGCTTTGAAAAG 68  
QY 61 CGCCGGCACTCCGCGATTTGATCCGCGCTTATGTCCTGTTAGGGCGCTCAGAGTCCGCG 120  
Db 69 CGCCGGCACTCCGCGATTTGATCCGCGCTTATGTCCTGTTAGGGCGCTCAGAGTCCGCG 128  
QY 121 TGGAGGCCATCACTCGCGCTTGGGAAACGCTCGAGCCCTACAGATTCAGAGTCA 180  
Db 129 TGGAGGCCATCACTCGCGCTTGGGAAACGCTCGAGCCCTACAGATTCAGAGTCA 188  
QY 181 TTTAAAGATATCAGATGGAGATTTGGGAAAAGGCAATTCAGACAGTCTTAAATGCT 240  
Db 189 TTTAAAGATATCAGATGGAGATTTGGGAAAAGGCAATTCAGACAGTCTTAAATGCT 248  
QY 241 GCAAAGGCTCTCAGATATGGCCCACTGATGAAAAGAGACATGTTGGCATGCTCATGCA 300  
Db 249 GCAAAGGCTCTCAGATATGGCCCACTGATGAAAAGAGACATGTTGGCATGCTCATGCA 308  
QY 301 GGAGAGGAGCTCCACACAGACCCAAAGAGGTTAAAGATGTTGATCTCGGAAAATC 360  
Db 309 GGAGAGGAGCTCCACACAGACCCAAAGAGGTTAAAGATGTTGATCTCGGAAAATC 368  
QY 361 ATAAAAGCAATGCTTCTTATGTGTGGCCCAAGAAGAGCCAGATCTACAGACTAGAGT 420  
Db 369 ATAAAAGCAATGCTTCTTATGTGTGGCCCAAGAAGAGCCAGATCTACAGACTAGAGT 428  
QY 421 GCCATTTGGCTGGATTTTGGGTGTGTGCAAAAGGCAATAATGTTGTTCCCTTCATG 480  
Db 429 CCCATTTGGCTGGATTTTGGGTGTGTGCAAAAGGCAATAATGTTGTTCCCTTCATG 488

QY 481 TTTAAATGCTGTAGACAGCCCTCAACAGATGTGGGAAAATGCTGAACTGAGTAT 540  
Db 489 TTTAAATGCTGTAGACAGCCCTCAACAGATGTGGGAAAATGCTGAACTGAGTAT 548  
QY 541 GCACCAATATAGTTGACACCATGACCAAGCAAGCACTTCTGATGTGCTATGCTACAGA 600  
Db 549 GCACCAATATAGTTGACACCATGACCAAGCAAGCACTTCTGATGTGCTATGCTATCAGA 608  
QY 601 GCTGAGCTGCTTTTAAAGAAAGTTGAAATGCAATGATTTGGCAAGTACCCAGAT 660  
Db 609 GCTGAGCTGCTTTTAAAGAAAGTTGAAATGCAATGATTTGGCAAGTACCCAGAT 668  
QY 661 TCAATCCGAGAAATAGCCAAAATGCTTTTCTCATCTTCAACACTGAGATCGGATTT 720  
Db 669 TCAATCCGAGAAATAGCCAAAATGCTTTTCTCATCTTCAACACTGAGATCGGATTT 728  
QY 721 CACCTGAGACAGACGCGAGCTTTATCTAAGCTATTTGACAGAGAACAGGGGTATC 780  
Db 729 CACCTGAGACAGACGCGAGCTTTATCTAAGCTATTTGACAGAGAACAGGGGTATC 788  
QY 781 AGTTTGTCCGAGGCTTGTGTATTTAATCTTCCCATCATGTTGAAGTATG--- 837  
Db 789 AGTTTGTCCGAGGCTTGTGTATTTAATCTTCCCATCATGTTGAAGTATGCTT 848  
QY 838 CTGTGCACTGCTGTTTGTATTAACAAATGCGGTCGCCAGTTGCTTTGTAACCTTGA 897  
Db 849 CTGTGCACTGCTGTTTGTATTAACAAATGCGTCCAGTT---GCTTGTACCTTGA 905  
QY 898 ACACTTGTATATACACAGCATTCACAGTTGCACTGACACGCTGAGAACATGATTTAGA 957  
Db 906 ACACTTGTATATACACAGCATTCACAGTTGCACTGACACGCTGAGAACATGATTTAGA 965  
QY 958 ATAGAAATGAACAAAGCAGATATGATGACAGTATGCTGCTATAGACTCATGCTGAT 1017  
Db 966 ATAGAAATGAACAAAGCAGATATGATGACAGTATGCTGCTATAGACTCATGCTGAT 1025  
QY 1018 TATGAACTGTGAAGATATTTAATATGAAGATATGAAAGACAGAGATATGATGATTT 1077  
Db 1026 TATGAACTGTGAAGATATTTAATATGAAGATATGAAAGACAGAGATATGATGATTT 1085  
QY 1078 TTGAAGACGTATGAGACTGCTTCACTGAAAAGTACCTTCACTGCGCTATCTGAACCTT 1137  
Db 1086 TTGAAGACGTATGAGACTGCTTCACTGAAAAGTACCTTCACTGCGCTATCTGAACCTT 1145  
QY 1138 GGTCAAGTGTATTTTACAGTGTGTTTAAACCTAATATGATGCTGCCAGTACAGGA 1197  
Db 1146 GGTCAAGTGTATTTTACAGTGTGTTTAAACCTAATATGATGCTGCCAGTACAGGA 1205  
QY 1198 ATTTGGCAGGTACCTTACTGTTGAGATCTAGTAATGTGAATGCACTCTTTTTCAG 1257  
Db 1206 ATTTGGCAGGTACCTTACTGTTGAGATCTAGTAATGTGAATGCACTCTTTTTCAG 1265  
QY 1258 CTTTCATTAACCCCTGAACCTTTCTGGGAACTGTATATAGAGACTAGACAACTACTA 1317  
Db 1266 CTTTCATTAACCCCTGAACCTTTCTGGGAACTGTATATAGAGACTAGACAACTACTA 1325  
QY 1318 GATATGAACACCTTGTATCTACTCAAGGTAGACCCAAATTAAGAACAAAGTATG 1377  
Db 1326 GATATGAACACCTTGTATCTACTCAAGGTAGACCCAAATTAAGAACAAAGTATG 1385  
QY 1378 GCATCTCCCTTACAGATCACACAGACAGCACTACCGTGCCTTTGATATATGCTATTT 1437  
Db 1386 GCATCTCCCTTACAGATCACACAGACAGCACTACCGTGCCTTTGATATATGCTATTT 1445  
QY 1438 GAATACATTTAGAGGCGCAAAAGTCTTATGTGAATATCTTTGAAGTCTCTGACGAAAG 1497  
Db 1446 GAATACATTTAGAGGCGCAAAAGTCTTATGTGAATATCTTTGAAGTCTCTGACGAAAG 1505  
QY 1498 AAGTGGCATTGTATGAGAGTGTGGGTCAAGGAAAGCAAAATGAGGCTATTTATTT 1557  
Db 1506 AAGTGGCATTGTATGAGAGTGTGGGTCAAGGAAAGCAAAATGAGGCTATTTATTT 1565

```

Qy 1558 CGCTTCTATGAGCTCAAAAGGGTAGCATTTATCTGCTGCTCAAAATATACAAAGATG 1617
Db 1566 CGCTTCTATGAGCTCAAAAGGGTAGCATTTATCTGCTGCTCAAAATATACAAAGATG 1625
Qy 1618 AGCCGGAAGACCTTGGAGGGAGAGTGGAGTGGTACCTGAGATGCTGCTCTTCAT 1677
Db 1626 AGCCGGAAGACCTTGGAGGGAGAGTGGAGTGGTACCTGAGATGCTGCTCTTCAT 1685
Qy 1678 AATACTATTTATTAACAACCTTTATATGAAACATACAGTGTCTTCACTGAGAGTGTAT 1737
Db 1686 AATACTATTTATTAACAACCTTTATATGAAACATACAGTGTCTTCACTGAGAGTGTAT 1745
Qy 1738 GCAAGTGGCAAAATTAAGTGGACCTTCAATGATCAATCTTCAATGCCATGATATGAC 1797
Db 1746 GCAAGTGGCAAAATTAAGTGGACCTTCAATGATCAATCTTCAATGCCATGATATGAC 1805
Qy 1798 ACCAAGTAGGGGGAAGAGAGCTCAAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGT 1857
Db 1806 ACCAAGTAGGGGGAAGAGAGCTCAAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGT 1865
Qy 1858 GCAAGAGCATTTTGAAGAGACCCCGAGTCACTACTATGATGAAGTCACTTCACTGTTA 1917
Db 1866 GCAAGAGCATTTTGAAGAGACCCCGAGTCACTACTATGATGAAGTCACTTCACTGTTA 1925
Qy 1918 GATTGATTAAGTGAAGAGACTATTTCTGTGTCATGAGAGATGTGTCAACACAGAACT 1977
Db 1926 GATTGATTAAGTGAAGAGACTATTTCTGTGTCATGAGAGATGTGTCAACACAGAACT 1985
Qy 1978 TCTATTTTCAATGACACAGATTTGTCAACAGTGGTGTATGAGATGAAATCATGTCTTG 2037
Db 1986 TCTATTTTCAATGACACAGATTTGTCAACAGTGGTGTATGAGATGAAATCATGTCTTG 2045
Qy 2038 GATCAGGGTAAAGTAGAGCGAAGCTGTATCCAGCAGTGTGTGCTTGAACCTCATAGT 2097
Db 2046 GATCAGGGTAAAGTAGAGCGAAGCTGTATCCAGCAGTGTGTGCTTGAACCTCATAGT 2105
Qy 2098 ATCTATTCAAGAAATGTGGCATACACAGAGCAGCCGTGTGCAAAACATGATACCCCAA 2157
Db 2106 ATCTATTCAAGAAATGTGGCATACACAGAGCAGCCGTGTGCAAAACATGATACCCCAA 2165
Qy 2158 TGGGAACCAAGAAAGAAATATATCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2217
Db 2166 TGGGAACCAAGAAAGAAATATATCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2225
Qy 2218 ATTGTCAATAGTGAAGAGCTGTGAAGACTGTGCTAAGTCAATAGACATTTTC 2277
Db 2226 ATTGTCAATAGTGAAGAGCTGTGAAGACTGTGCTAAGTCAATAGACATTTTC 2285
Qy 2278 TTTTGTGTGTGTGTGACATATATTTGCACTGAGCAGAAATGTTTATTTAAAAAAT 2337
Db 2286 TTTTGTGTGTGTGTGACATATATTTGCACTGAGCAGAAATGTTTATTTAAAAAAT 2345
Qy 2338 CATACATT 2345
Db 2346 CATACATT 2353

RESULT 2
LOCUS U66679 1066 bp mRNA linear HTC 23-JUL-2001
DEFINITION Homo sapiens clone EST140535 mRNA sequence.
ACCESSION U66679
VERSION U66679.1 GI:1906564
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1066)
AUTHORS Allikmets,R., Gerrard,B., Hutchinson,A. and Dean,M.
TITLE Characterization of the human ABC superfamily: isolation and
mapping of 21 new genes using the expressed sequence tags database
JOURNAL Hum. Mol. Genet. 5 (10), 1649-1655 (1996)

```

```

MEDLINE 97049974
PUBMED 8894702
REFERENCE 2 (bases 1 to 1066)
AUTHORS Allikmets,R., Gerrard,B. and Dean,M.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1996) Human Genetics Section, National Cancer
Institute, NCI-FCRDC, Frederick, MD 21702, USA
FEATURES
source
1. 1066
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="EST140535"
/note="similar to ATP-binding cassette transporter"
BASE COUNT 342 a 201 c 242 g 281 t
ORIGIN
Query Match 45.3%; Score 1061.2; DB 11; Length 1066;
Best Local Similarity 99.7%; Pred. No. 1,6e-23; Indels 0; Gaps 0;
Matches 1063; Conservative 0; Mismatches 3;
Qy 1280 TGGGAAGTATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1339
Db 1 TGGGAAGTATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Qy 1340 TACTCAAGTAGACACCCCAATTAAGACAAAGTAGATGAGATCTCCCTTCAGATCAC 1399
Db 61 TACTCAAGTAGACACCCCAATTAAGACAAAGTAGATGAGATCTCCCTTCAGATCAC 120
Qy 1400 CACAGACAGTACCGTGGCTTTGATATGTGATTTGAAATACATTAGAGGCCAGAAAG 1459
Db 121 CACAGACAGTACCGTGGCTTTGATATGTGATTTGAAATACATTAGAGGCCAGAAAG 180
Qy 1460 TCCCTTAGGAAATATCCTTTGAAGTCCCTGACAGAAAGAGAGAGAGAGAGAGAG 1519
Db 181 TCCCTTAGGAAATATCCTTTGAAGTCCCTGACAGAAAGAGAGAGAGAGAGAGAG 240
Qy 1520 GTGGGTGAGGAAAGACAAATAGTAGAGCTATATTTCGCTTCTATGAGCCTCAAAAG 1579
Db 241 GTGGGTGAGGAAAGACAAATAGTAGAGCTATATTTCGCTTCTATGAGCCTCAAAAG 300
Qy 1580 GTAGCATTTATCTTGTGCTGCTCAAAATATACAGATGTAGCCTGGAAGCCTTGGAGGG 1639
Db 301 GTAGCATTTATCTTGTGCTGCTCAAAATATACAGATGTAGCCTGGAAGCCTTGGAGGG 360
Qy 1640 CAGTGGAGAGTGAACCTCAGAGAGTGTCTGCTCTCATATATACATTTATTAACAACCT 1699
Db 361 CAGTGGAGAGTGAACCTCAGAGAGTGTCTGCTCTCATATATACATTTATTAACAACCT 420
Qy 1700 TATATGAAACATCAGTGTCTTCACTGAGAGAGTATGAGAGAGAGAGAGAGAGAG 1759
Db 421 TATATGAAACATCAGTGTCTTCACTGAGAGAGTATGAGAGAGAGAGAGAGAGAG 480
Qy 1760 TTATATGATGAATTTTCCATGATGCCATGATATGACACCCAAAGTAGGGGAAAGAG 1819
Db 481 TTATATGATGAATTTTCCATGATGCCATGATATGACACCCAAAGTAGGGGAAAGAG 540
Qy 1820 TCAAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1879
Db 541 TCAAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy 1880 CCCAGTCACTCTATGATGAAGTGAAGTCACTGTTAGATGATGATGATGATGATGAT 1939
Db 601 CCCAGTCACTCTATGATGAAGTGAAGTCACTGTTAGATGATGATGATGATGATGAT 660
Qy 1940 TTCTTGTGTCATGAAGAGATGTGTCAAAACACAGAACTTCTATTTTCATGACACAGAT 1999
Db 661 TTCTTGTGTCATGAAGAGATGTGTCAAAACACAGAACTTCTATTTTCATGACACAGAT 720
Qy 2000 TGTCAACAGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2059
Db 721 TGTCAACAGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy 2060 GTGTACCCACCATGTTGCTTGTCTTACCTCATAGTATCTATGAGAAATGTGACATA 2119

```

Db 73 CCAGATTCAATCCGAGAATAGCCAAATGTCTTTCTCCATCTTCAACAACCTGGATCT 132

|         |  |
|---------|--|
| AUTHORS | Bonaldo, M. F., Lennon, G. and Soares, M. B.   |
| TITLE   | Normalization and subtraction: two approaches to facilitate gene expression analysis |

JOURNAL  
Genome Res. 6 (9), 791-806 (1996)  
MEDLINE  
97044477  
COMMENT  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.

# FEATURES

Location/Qualifiers

1..730  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-E-EO1-alc-o-22-0-UI"  
/clone\_1lb="UI-E-EO1"  
/cissue\_type="fetal eye"  
/dev\_stage="fetal"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: eye; Vector: pRTT3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-EO1 is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I and cloned directionally into pRTT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGCGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 213 a 142 c 163 g 211 t 1 others

ORIGIN

Query Match 29.9%; Score 700.4; DB 14; Length 730;  
Best Local Similarity 98.2%; Pred. No. 5.7e-144;  
Matches 718; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 997 GCTATAGACTCACTGCTGATTAATGAAGCTGGAATTTTATATGAAGATATGAA 1056

Db 1 GCTATAGACTCACTGCTGATTAATGAAGCTGGAATTTTATATGAAGATATGAA 60

QY 1057 GCACAGAGATATGATGATTTTGAAGAGCTATGAGCTCTTCATTTGAAAGTACCTCT 1116

Db 61 GCACAGAGATATGATGATTTTGAAGAGCTATGAGCTCTTCATTTGAAAGTACCTCT 120

QY 1117 ACTGTGCTATGCTGAATTTTGTCAAGTCTATTTTCACTGTCGCTTTAAACAGCTATA 1176

Db 121 ACTGTGCTATGCTGAATTTTGTCAAGTCTATTTTCACTGTCGCTTTAAACAGCTATA 180

QY 1177 ATGTGTCTCCAGTCAAGGAATTTGTGAGAGTACCTTACTGTGAGATCTAGTAATG 1236

Db 181 ATGTGTCTCCAGTCAAGGAATTTGTGAGAGTACCTTACTGTGAGATCTAGTAATG 240

QY 1237 GTGAATGAGTCTTTTTCAGCTTTTCACTTACCCCTGAACCTTTCGGAACTGTATATGA 1296

Db 241 GTGAATGAGTCTTTTTCAGCTTTTCACTTACCCCTGAACCTTTCGGAACTGTATATGA 300

QY 1297 GAGACTGACAAGACATATGATATGAACCTTGTTTACTCTACAGATGACACACC 1356

Db 301 GAGACTGACAAGACATATGATATGAACCTTGTTTACTCTACAGATGACACACC 360

QY 1357 CAATTAAGACAAAGTATGATGATCTCCCTTCAGATCAACACAGACAGCTACCGTG 1416

Db 361 CAATTAAGACAAAGTATGATGATCTCCCTTCAGATCAACACAGACAGCTACCGTG 420

QY 1417 GCCTTTGATATGATGATTTTGAATACATTTAGAGGCGCAAAATCTCTATGAAATTC 1476

Db 421 GCCTTTGATATGATGATTTTGAATACATTTAGAGGCGCAAAATCTCTATGAAATTC 480

QY 1477 TTTGAAGTCCCTGACGAAAGAAAGTGGCATTGTAGAGGTAGTGGTCAAGGAAAGC 1536

Db 481 TTTGAAGTCCCTGACGAAAGAAAGTGGCATTGTAGAGGTAGTGGTCAAGGAAAGC 540

QY 1537 ACAATATGAGAGCTATTTATTTGCTTCTATGAGCCTCAAAAGGATGATTTATCTTGT 1596

Db 541 ACAATATGAGAGCTATTTATTTGCTTCTATGAGCCTCAAAAGGATGATTTATCTTGT 600

QY 1597 GGTCAAAATATACAAATGTGAGACCTTGAAAGCCTTCGAGGCGAGTGGAGTGTACT 1656

Db 601 GGTCAAAATATACAAATGTGAGACCTTGAAAGCCTTCGAGGCGAGTGGAGTGTACT 659

QY 1657 CAGATGCTGTCTCTTCATATATCTATTTATCAACCTTTATATGGAACATCAGT 1716

Db 660 CAGATGCTGTCTCTTCATATATCTATTTATCAACCTTTATATGGAACATCAGT 719

QY 1717 GCTTCACTGA 1727

Db 720 GCTTCACTGA 730

RESULT 5

LOCUS BG722394 787 bp mRNA EST 08-MAY-2001

DEFINITION 602693606F1 NIH\_MGC\_97 Homo sapiens CDNA clone IMAGE:4825873 5',

ACCESSION BG722394

VERSION BG722394.1 GI:14001581

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM10739 row: 1 column: 02

High quality sequence stop: 771.

## FEATURES

source

1..787

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="MIMR:4825873"

/clone\_1lb="NIH\_MGC\_97"

/lab\_host="DH10B"

/note="Organ: testis; Vector: pBluescript (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gicgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI/NHGR), National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 216 a 175 c 211 g 185 t

ORIGIN

Query Match 29.7%; Score 696.6; DB 12; Length 787;  
Best Local Similarity 97.4%; Pred. No. 3.9e-143;  
Matches 751; Conservative 0; Mismatches 14; Indels 6; Gaps 4;

1 ATGGCGCTGCTCGCGATCATCTTCTGGCGCTGGCGCGCGCGCTGCTCTTCCGAAAG 60  
17 ATGGCGCTGCTCGCGATCATCTTCTGGCGCTGGCGCGCGCGCTGCTCTTCCGAAAG 76  
61 CGCGGCACTCCGCGATCTGATCCGCGCTTATCTCTTTAGCGCTCAGGTCCGAG 120  
77 CGCGGCACTCCGCGATCTGATCCGCGCTTATCTCTTTAGCGCTCAGGTCCGAG 136  
121 TGGAGGCCACATCACTGGCGGCTTGGGAAACGCTCGAGGCTAC---CAGATTCAGAG 177  
137 TGGAGGCCACATCACTGGCGGCTTGGGAAACGCTCGAGGCTACGAGATTCGAGAG 196  
178 TCATTAATAAGATATCATATGCGAGAGATTGGGAAAGCAATTCAGACAGTCTTAGAT 237  
197 TCATTAATAAGATATCATATGCGAGAGATTGGGAAAGCAATTCAGACAGTCTTAGAT 256  
238 GCTGCAAAAGCTCTCCAGGATATGCCACTGATGAAAAGAGAGACATGTTGCGATGCTAT 297  
257 GCTGCAAAAGCTCTCCAGGATATGCCACTGATGAAAAGAGAGACATGTTGCGATGCTAT 316  
298 GCAGAGAGAGAGCTCCACACAGACCCCAAAAGAGGTTTAAAGATGTTGATCTCGGAAA 357  
317 GCAGAGAGAGAGCTCCACACAGACCCCAAAAGAGGTTTAAAGATGTTGATCTCGGAAA 376  
358 ATCATTAATAAGATATCATATGCTTATGTTGGGCCCAAGACAGGCGCATCTAGAGCTAGA 417  
377 ATCATTAATAAGATATCATATGCTTATGTTGGGCCCAAGACAGGCGCATCTAGAGCTAGA 436  
418 GTTGCCATTTGCTGGGATTTTGGGTGGTGCAAAAGCCATGATATTTGTTCCCTTC 477  
437 GTTGCCATTTGCTGGGATTTTGGGTGGTGCAAAAGCCATGATATTTGTTCCCTTC 495  
478 ATGTTAAATATGCTGTGAGACAGGCTCAACAGATGTGGGAAACATGCTGAACCTGAGT 537  
496 ATGTTAAATATGCTGTGAGACAGGCTCAACAGATGTGGGAAACATGCTGAACCTGAGT 554  
538 GATGCAACCAATATACAGTTGCAACCATGCGAACAGAGTCTGATGGCTATGCTATCA 597  
555 GATGCAACCAATATACAGTTGCAACCATGCGAACAGAGTCTGATGGCTATGCTATCA 614  
598 AGAGCTGAGCTGCTTTTAAAGAGTTGCAAAATGCAATGTTGGCAAGGTAGCCGAG 657  
615 AGAGCTGAGCTGCTTTTAAAGAGTTGCAAAATGCAATGTTGGCAAGGTAGCCGAG 674  
658 AATTCAATCCCAAGAAATAGCCCAAAATGCTTCTCCATCTTCAACACCTGATCTGGGT 717  
675 AATTCAATCCCAAGAAATAGCAAAAAATGCTTCTTACATCTTCAACA-CTGATCTGGGT 733  
718 TTTCACTGAGCAGACAGACGGAGCTTATCTAAGGCTATGACAGAGA 768  
734 TTTCACTGAGCAGACAGACGGAGCTTATCTAAGGCTATGACAGAGA 784

RESULT 6  
BI463667  
LOCUS 603207337P1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5273168 5',  
DEFINITION mRNA sequence.  
ACCESSION BI463667  
VERSION BI463667.1 GI:15254323  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 892)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shizaki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.lnl.gov>  
Plate: L1M1689 row: 1 column: 09  
High quality sequence stop: 721.  
Location/Qualifiers

FEATURES  
source  
1.892  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5273168"  
/clone\_1ib="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.2 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 231 a 201 c 237 g 223 t

ORIGIN

Query Match 29.2%; Score 685.8; DB 13; Length 892;  
Best Local Similarity 95.0%; Pred. No. 9.3e-141;  
Matches 829; Conservative 0; Mismatches 27; Indels 17; Gaps 11;

1 ATGGCGCTGCTCGCGATCATCTTGGCGCTGGCGCGCGCGCTGCTCTTCCGAAAG 60  
17 ATGGCGCTGCTCGCGATCATCTTGGCGCTGGCGCGCGCGCTGCTCTTCCGAAAG 76  
61 CGCGGCACTCCGCGATCTGATCCGCGCTTATCTCTGTTAGCGGCTCAGTCCGAG 120  
77 CGCGGCACTCCGCGATCTGATCCGCGCTTATCTCTGTTAGCGGCTCAGTCCGAG 136  
121 TGGAGGCCACATCACTCGGCGCTTGGGAAACGCTCGAGCTAC---CAGATTCAGAG 177  
137 TGGAGGCCACATCACTCGGCGCTTGGGAAACGCTCGAGCTACGAGATTCAGAG 196  
178 TCATTAATAAGATATCATATGCGAGAGATTGGGAAAGCAATTCAGACAGTCTTAGAT 237  
197 TCATTAATAAGATATCATATGCGAGAGATTGGGAAAGCAATTCAGACAGTCTTAGAT 256  
238 GCTGCAAAAGCTCTCCAGGATATGCCACTGATGAAAAGAGAGATGTTGCGATGCTAT 297  
257 GCTGCAAAAGCTCTCCAGGATATGCCACTGATGAAAAGAGAGATGTTGCGATGCTAT 316  
298 GCAGAGAGAGAGCTCCACACAGACCCCAAAAGAGGTTTAA--AGATCTGATCTCGGA 355  
317 GCAGAGAGAGAGCTCCACACAGACCCCAAAAGAGGTTTAA--AGATCTGATCTCGGA 376  
356 AATCATTAATAAGCAATGCTTATGTTGGGCCCAAGAGACAGGCTAGTCAAGAGTA 415  
377 AATCATTAATAAGCAATGCTTATGTTGGGCCCAAGAGACAGGCTAGTCAAGAGTA 436  
416 GAGTTGCCATTTGCTGGGATTTTGGGTGGTGCAAAAGCCATGATATTTGCTTCCT 475  
437 GAGTTGCCATTTGCTGGGATTTTGGGTGGTGCAAAAGCCATGATATTTGCTTCCT 495  
476 TCATGTTAAATATGCTGTAGACAGCTCAACCAAGATGTGGGAAACATGCTGAACCTGA 535  
496 TCATGTTAAATATGCTGTAGACAGCTCAACCAAGATGTGGGAAACATGCTGAACCTGA 555

QY 536 GTGATGCACCAATACAGTTGGCAACCATGGCAACAGCAGTTCTGATTGGCTAGTGTAT 595  
 |||||  
 Db 556 GTGATGCACCAATACAGTTGGCAACCATGGCAACAGCAGTTCTGATTGGCTAGTGTAT 615  
 |||||  
 QY 596 CAAGAGCTGGAGCTGCTTTTAAACGAAGTTGCAAGTATGAGTATTTGGCAAGTAG-CC 654  
 |||||  
 Db 616 CAAGAGCTGGAGCTGCTTTTAAACGAAGTTGCAAGTATGAGTATTTGGCAAGTAGCCCC 675  
 |||||  
 QY 655 CAGATTCAATCCGAGAGATA-GCCAAAATGTCTT-CTCCATCTTCAACAACCTGG-- 709  
 |||||  
 Db 676 CAGATTCAATCCGAGAGATAATGCCAAAATGTCTTACTCTTCAACAACCTGGCAT 735  
 |||||  
 QY 710 ATCTGGGTTTACCTGAGCAGACA-GACGGAGCTTTATCT-AAAGCTATTGACAGAG 767  
 |||||  
 Db 736 CTGGGGTTTACACCTGAGCAGACACGCGAGCTTTATCTAAAGCTATTGACAGAG 795  
 |||||  
 QY 768 AACAA--GGGTATCACTTTTGTCTGAGTCTTGGTATTTAATCTTCTCCATCATG 825  
 |||||  
 Db 796 GACCACGCGGTATCACTTTGTCTGAGTCTGCGGTATTAATCTTCTCCATCATG 855  
 |||||  
 QY 826 TTGAAGTGAAGCTTGTCACTGAGTGTGTGTAT 858  
 |||||  
 Db 856 -TTGAATGATGCTTGTCAAGGGGTTGATTT 887  
 |||||

RESULT 7  
 AU127740 704 bp mRNA linear EST 01-AUG-2002  
 LOCUS AU127740 NT2RP2 Homo sapiens cDNA clone NT2RP2001958 5', mRNA  
 DEFINITION sequence.

ACCESSION AU127740  
 VERSION AU127740  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 704)  
 Oca.T., Nishikawa.T., Suzuki.Y., Iehi.S., Saito.K., Kawai.Y.,  
 Yamamoto.J., Makamatsu.A., Nakamura.Y., Nagai.T., Sugano.S. and  
 Isogai.T.

TITLE HRI human cDNA project  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3975  
 Fax: 81-438-52-3986  
 Email: genomics@hri.co.jp

HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix  
 Research Institute; cDNA library construction: Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 Location/Qualifiers  
 1. 704  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP2001958"  
 /clone\_lib="NT2RP2"  
 /cell\_type="teratocarcinoma"  
 /cell\_line="NT2"  
 /note="Vector: pME18SFL3; RNA from NT2 neuronal precursor  
 cells after 2-weeks retinoic acid (RA) induction."

## FEATURES

source

BASE COUNT 190 a 142 c 168 g 198 t  
 ORIGIN  
 Query Match 27.9%; Score 655.2; DB 9; Length 704;  
 Best Local Similarity 97.0%; Pred. No. 5.3e-134;  
 Matches 665; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

QY 232 TTAGATGCTCAAAAGGCTCTCCAGTATGGCCACTGATAGAAAAAGAGACATGTGGCAT 291

Db 1 TTAGATGCTCAAAAGGCTCTCCAGTATGGCCACTGATAGAAAAAGAGACATGTGGCAT 60  
 |||||  
 QY 292 GGTATGTCAGAGAGAGAACTCCACAGACCCCAAAAGAGGTTAAAGATGTTGTAAT 351  
 |||||  
 Db 61 GGTATGTCAGAGAGAGAACTCCACAGACCCCAAAAGAGGTTAAAGATGTTGTAAT 120  
 |||||  
 QY 352 CGGAAATATCAATAAGCAATGCTTTCTTATGTGTGGCCCAAGACAGCCGAGTCTACGA 411  
 |||||  
 Db 121 CGGAAATATCAATAAGCAATGCTTTCTTATGTGTGGCCCAAGACAGCCGAGTCTACGA 180  
 |||||  
 QY 412 GCTAGAGTTGCCATTTTGGCTGGATTTTGGGTGTGTGCAAGGCCATGATATTGTGTT 471  
 |||||  
 Db 181 GCTAGAGTTGCCATTTTGGCTGGATTTTGGGTGTGTGCAAGGCCATGATATTGTGTT 240  
 |||||  
 QY 472 CCTTATGTTTAAATATGCTGTAGACAGCTTCAACAGATGTGGGAAACATGCTGAAC 521  
 |||||  
 Db 241 CCTTATGTTTAAATATGCTGTAGACAGCTTCAACAGATGTGGGAAACATGCTGAAC 300  
 |||||  
 QY 532 CTGAGTATGACCAATAATACAGTTGCAACCATGGCAACAGCAGTTGATTTGCTATGCT 591  
 |||||  
 Db 301 CTGAGTATGACCAATAATACAGTTGCAACCATGGCAACAGCAGTTGATTTGCTATGCT 360  
 |||||  
 QY 592 GTATCAAGAGCTGGAGCTGTTTTTTAAACGAAGTTGCAATGCAATTTGGCAAGTGA 651  
 |||||  
 Db 361 GTATCAAGAGCTGGAGCTGTTTTTTAAACGAAGTTGCAATGCAATTTGGCAAGTGA 420  
 |||||  
 QY 652 GCCCAGAAATTCATCCGAGAAATAGCCAAATATGCTTCTCATCTTCAACACCTGGAT 711  
 |||||  
 Db 421 GCCCAGAAATTCATCCGAGAAATAGCCAAATATGCTTCTCATCTTCAACACCTGGAT 480  
 |||||  
 QY 712 CTGGGTTTTCACTGAGCAGACAGACGAGAGCTTTATCTAAGGCTATTGACAGAGAGA 771  
 |||||  
 Db 481 CTGGGTTTTCTGAGCAGACAGACGAGAGCTTTATCTAAGGCTATTGACAGAGAGA 539  
 |||||  
 QY 772 AGGGGATATCAATTTTGTCTGAGTCTTGTGATTTAATCTTCTCCATCATGTTGAA 831  
 |||||  
 Db 540 AGGGGATATCAATTTTGTCTGAGTCTTGTGATTTAATCTTCTCCATCATGTTGAA 598  
 |||||  
 QY 832 GTGATGCTGTACAGAGGCTTTTGTATTAACAATGGGCGCCAGTTGCTTGGTAAC 891  
 |||||  
 Db 599 GTGATGCTGTACAGAGGCTTTTGTATTAACAATGGGCGCCAGTTGCTTGGTAAC 658  
 |||||  
 QY 892 CTGGAACACTTGATATACACAGCATTCACAGTTGACGTACAC 937  
 |||||  
 Db 659 CTTGGAACACTTGATATACACAGCATTCACAGTTGACGTACAC 704  
 |||||

RESULT 8  
 BE876927 822 bp mRNA linear EST 20-OCT-2000  
 LOCUS BE876927 601484702P1 NIH\_MGC\_69 Homo sapiens cDNA clone IMAGE:3887161 5',  
 mRNA sequence.

ACCESSION BE876927  
 VERSION BE876927  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 822)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

## AUTHORS

TITLE  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov

Tissue Procurement: DCTD/DRP/Gazdar  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
 http://image.llnl.gov







| QY  | 300   | AGGAGAGGACCTCCACACGACCCCAAAAGAGTTTAAAGATGTGATCTACCTGGAAAT                     | 359                    |
|---|---|---|------------------------|
| QY <td>300 <td>AGGAGAGGACCTCCACACGACCCCAAAAGAGTTTAAAGATGTGATCTACCTGGAAAT <td>359</td> </td></td>        | 300 <td>AGGAGAGGACCTCCACACGACCCCAAAAGAGTTTAAAGATGTGATCTACCTGGAAAT <td>359</td> </td>        | AGGAGAGGACCTCCACACGACCCCAAAAGAGTTTAAAGATGTGATCTACCTGGAAAT <td>359</td>        | 359                    |
| Db <td>315 <td>AGGAGAGGACCTCCACACGACCCCAAAAGAGTTTAAAGATGTGATCTACCTGGAAAT <td>374</td> </td></td>        | 315 <td>AGGAGAGGACCTCCACACGACCCCAAAAGAGTTTAAAGATGTGATCTACCTGGAAAT <td>374</td> </td>        | AGGAGAGGACCTCCACACGACCCCAAAAGAGTTTAAAGATGTGATCTACCTGGAAAT <td>374</td>        | 374                    |
| QY <td>360 <td>CATTAAGCAATGCTTTCTTATGTGTGCGCCCAAGACAGCCAGATCTACGAGCTAGAGT <td>419</td> </td></td>       | 360 <td>CATTAAGCAATGCTTTCTTATGTGTGCGCCCAAGACAGCCAGATCTACGAGCTAGAGT <td>419</td> </td>       | CATTAAGCAATGCTTTCTTATGTGTGCGCCCAAGACAGCCAGATCTACGAGCTAGAGT <td>419</td>       | 419                    |
| Db <td>375 <td>CATTAAGCAATGCTTTCTTATGTGTGCGCCCAAGACAGCCAGATCTACGAGCTAGAGT <td>434</td> </td></td>       | 375 <td>CATTAAGCAATGCTTTCTTATGTGTGCGCCCAAGACAGCCAGATCTACGAGCTAGAGT <td>434</td> </td>       | CATTAAGCAATGCTTTCTTATGTGTGCGCCCAAGACAGCCAGATCTACGAGCTAGAGT <td>434</td>       | 434                    |
| QY <td>420 <td>TGCCATTTGCTGGAGATTTTGGGTGTGTGCAAGGCCATGAATATTTGTTCCTTCAT <td>479</td> </td></td>         | 420 <td>TGCCATTTGCTGGAGATTTTGGGTGTGTGCAAGGCCATGAATATTTGTTCCTTCAT <td>479</td> </td>         | TGCCATTTGCTGGAGATTTTGGGTGTGTGCAAGGCCATGAATATTTGTTCCTTCAT <td>479</td>         | 479                    |
| Db <td>435 <td>TGCCATTTGCTGGAGATTTTGGGTGTGTGCAAGGCCATGAATATTTGTTCCTTCAT <td>494</td> </td></td>         | 435 <td>TGCCATTTGCTGGAGATTTTGGGTGTGTGCAAGGCCATGAATATTTGTTCCTTCAT <td>494</td> </td>         | TGCCATTTGCTGGAGATTTTGGGTGTGTGCAAGGCCATGAATATTTGTTCCTTCAT <td>494</td>         | 494                    |
| QY <td>480 <td>GTTTAAATATGTGTGTGACAGCGCTCAACCAATGTCTGGGAAACATGCTGAACCTTAGTGA <td>539</td> </td></td>    | 480 <td>GTTTAAATATGTGTGTGACAGCGCTCAACCAATGTCTGGGAAACATGCTGAACCTTAGTGA <td>539</td> </td>    | GTTTAAATATGTGTGTGACAGCGCTCAACCAATGTCTGGGAAACATGCTGAACCTTAGTGA <td>539</td>    | 539                    |
| Db <td>495 <td>GTTTAAATATGTGTGTGACAGCGCTCAACCAATGTCTGGGAAACATGCTGAACCTTAGTGA <td>554</td> </td></td>    | 495 <td>GTTTAAATATGTGTGTGACAGCGCTCAACCAATGTCTGGGAAACATGCTGAACCTTAGTGA <td>554</td> </td>    | GTTTAAATATGTGTGTGACAGCGCTCAACCAATGTCTGGGAAACATGCTGAACCTTAGTGA <td>554</td>    | 554                    |
| QY <td>540 <td>TGCACCAAAATACAGTTGCAACCATGAGCAACAGACAGTTGCTGATTTGGCTATGAGTATCAAG <td>599</td> </td></td> | 540 <td>TGCACCAAAATACAGTTGCAACCATGAGCAACAGACAGTTGCTGATTTGGCTATGAGTATCAAG <td>599</td> </td> | TGCACCAAAATACAGTTGCAACCATGAGCAACAGACAGTTGCTGATTTGGCTATGAGTATCAAG <td>599</td> | 599                    |
| Db <td>555 <td>TGCACCAAAATACAGTTGCTGATTTGGCTATGAGTATCAAG <td>613</td> </td></td>                        | 555 <td>TGCACCAAAATACAGTTGCTGATTTGGCTATGAGTATCAAG <td>613</td> </td>                        | TGCACCAAAATACAGTTGCTGATTTGGCTATGAGTATCAAG <td>613</td>                        | 613                    |
| QY <td>600 <td>AGCTGAGCTGCTTTTAAAAAGCAAGTTGCAAGATGCAATTTTGGCAAGTACCCAGAA <td>659</td> </td></td>        | 600 <td>AGCTGAGCTGCTTTTAAAAAGCAAGTTGCAAGATGCAATTTTGGCAAGTACCCAGAA <td>659</td> </td>        | AGCTGAGCTGCTTTTAAAAAGCAAGTTGCAAGATGCAATTTTGGCAAGTACCCAGAA <td>659</td>        | 659                    |
| Db <td>614 <td>AGCTGAGCTGCTTTTAAAAAGCAAGTTGCAAGATGCAATTTTGGCAAGTACCCAGAA <td>673</td> </td></td>        | 614 <td>AGCTGAGCTGCTTTTAAAAAGCAAGTTGCAAGATGCAATTTTGGCAAGTACCCAGAA <td>673</td> </td>        | AGCTGAGCTGCTTTTAAAAAGCAAGTTGCAAGATGCAATTTTGGCAAGTACCCAGAA <td>673</td>        | 673                    |
| QY <td>660 <td>TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>717</td> </td></td>        | 660 <td>TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>717</td> </td>        | TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>717</td>        | 717                    |
| Db <td>674 <td>TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>732</td> </td></td>        | 674 <td>TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>732</td> </td>        | TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>732</td>        | 732                    |
| QY <td>718 <td>TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>734</td> </td></td>        | 718 <td>TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>734</td> </td>        | TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>734</td>        | 734                    |
| Db <td>733 <td>TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>749</td> </td></td>        | 733 <td>TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>749</td> </td>        | TTTCATTCGCAAAATATGCAAAATATGCTTTTCCATCTTCCAACTGAGATCTGG--T <td>749</td>        | 749                    |
| RESULT 10   | BM975299/c  |   |                        |
| LOCUS   | BM975299/c  | 666 bp  | linear EST 21-MAR-2002 |
| DEFINITION  | UI-CF-EC1-ack-e-24-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone                                |   |                        |
| ACCESSION   | BM975299  |   |                        |
| VERSION   | BM975299.1  | GI:19592890   |                        |
| KEYWORDS  | EST.  |   |                        |
| SOURCE  | human.  |   |                        |
| ORGANISM  | Homo sapiens  |   |                        |
| REFERENCE   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;                           |   |                        |
| AUTHORS   | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.                                  |   |                        |
| TITLE   | 1 (bases 1 to 666)  |   |                        |
| JOURNAL   | Bonaldi M.F., Lennon G. and Soares M.B.   |   |                        |
| MEDLINE   | Normalization and subtraction: two approaches to facilitate gene                            |   |                        |
| COMMENT   | discovery   |   |                        |
|   | Genome Res. 6 (9) , 791-806 (1996)  |   |                        |
|   | 97044477  |   |                        |
|   | Contact: McCray, PB   |   |                        |
|   | McCray Lab  |   |                        |
|   | University of Iowa  |   |                        |
|   | 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA                                  |   |                        |
|   | Tel: 319 356 4866   |   |                        |
|   | Fax: 319 356 7171   |   |                        |
|   | Email: paul-mccray@iowa.edu   |   |                        |
|   | Tissue Procurement: Dr. M. J. Welsh, University of Iowa                                     |   |                        |
|   | cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa                           |   |                        |
|   | cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa                            |   |                        |
|   | DNA Sequencing by: Dr. M. Bento Soares, University of Iowa                                  |   |                        |
|   | Clone Distribution: Researchers may obtain clones from Research                             |   |                        |
|   | Genetics (www.resgen.com).  |   |                        |
|   | Seq primer: M13 FORWARD   |   |                        |
|   | POLYA=yes.  |   |                        |
| FEATURES  | Location/Qualifiers   |   |                        |
| source  | 1..666  |   |                        |
|   | /organism="Homo sapiens"  |   |                        |
|   | /db_xref="taxon:9606"   |   |                        |
|   | /c_lone="UI-CF-EC1-ack-e-24-0-UI"   |   |                        |
|   | /clone_lib="UI-CF-EC1"  |   |                        |
|   | /feature_type="Lung"  |   |                        |

```

/dev_stg="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: p7713-Pac (Pharmacia) with a
modified polylinker; Site_1: SCor 1; Site_2: Not 1;
UI-CF-EC1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into p7713-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AAGGCTTAC.
TAG_LIB=UI-CF-EC1
TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 363-371
and 380-383
TAG_SEQ=AAGGCTTAC"
BASE COUNT      169 a      145 c      118 g      234 t
ORIGIN

```

| Query Match           | 27.5%  | Score 645          | DB 14    | Length 666 |
|-----------------------|--|--------------------|----------|------------|
| Best Local Similarity | 100.0%   | Pred. No. 9.4e-132 |          |            |
| Matches 645           | Conservative 0   | Mismatches 0       | Indels 0 | Gaps 0     |
| Qy 1701               | ATATGAAACATCAAGTCTTCACTGAGGAAGTGATGACAGTGGCAAAATTAGTGA | ACT                | 1760     |            |
| Db 666                | ATATGAAACATCAAGTCTTCACTGAGGAAGTGATGACAGTGGCAAAATTAGTGA | ACT                | 607      |            |
| Qy 1761               | TCATATGCAATTCTTCGAATGCCACATGATATGACACCCAAATGAGGGAACGAG | ACT                | 1820     |            |
| Db 606                | TCATATGCAATTCTTCGAATGCCACATGATATGACACCCAAATGAGGGAACGAG | ACT                | 547      |            |
| Qy 1821               | CAAGCTTTCAGAGGAGAAAGCAAAAGATGCAATTGCAAGCCATTTTGAAGAC   | CCC                | 1880     |            |
| Db 546                | CAAGCTTTCAGAGGAGAAAGCAAAAGATGCAATTGCAAGCCATTTTGAAGAC   | CCC                | 487      |            |
| Qy 1881               | CCCACTACTCTATGATGAAGCTACTTCAATCGTTAGATTCGATTCTGAAGAG   | CTAT               | 1940     |            |
| Db 486                | CCCACTACTCTATGATGAAGCTACTTCAATCGTTAGATTCGATTCTGAAGAG   | CTAT               | 427      |            |
| Qy 1941               | TCTTGATCCATGAAAGATGATGTCAAACAGAACTTCTATTTTCATTTGCACAG  | ATT                | 2000     |            |
| Db 426                | TCTTGATCCATGAAAGATGATGTCAAACAGAACTTCTATTTTCATTTGCACAG  | ATT                | 367      |            |
| Qy 2001               | GTCAAACAGTGGTTGATGCAGATGAATCATTTGTTGGATCAGGGTAAAGTA    | CCG                | 2060     |            |
| Db 366                | GTCAAACAGTGGTTGATGCAGATGAATCATTTGTTGGATCAGGGTAAAGTA    | CCG                | 307      |            |
| Qy 2061               | TGTATCCCAACCATGGTTTGCTTGCTAACCCCTCATATGATCTATTCAGAA    | ATGTGGCATAC        | 2120     |            |
| Db 306                | TGTATCCCAACCATGGTTTGCTTGCTAACCCCTCATATGATCTATTCAGAA    | ATGTGGCATAC        | 247      |            |
| Qy 2121               | ACAGAGCAGCCGTGTGCAGAACCATGATTAACCCCAATGAGAGCAAAAGAA    | ATAAT              | 2180     |            |
| Db 246                | ACAGAGCAGCCGTGTGCAGAACCATGATTAACCCCAATGAGAGCAAAAGAA    | ATAAT              | 187      |            |
| Qy 2181               | ATCCAAAGAGAGGAAGAAAGAACTTCAAGAAATAATGTCATATGATGTGA     | AGGCTG             | 2240     |            |
| Db 186                | ATCCAAAGAGAGGAAGAAAGAACTTCAAGAAATAATGTCATATGATGTGA     | AGGCTG             | 127      |            |
| Qy 2241               | TGGAACCTGTCGTCTAATGACATATAGCATTTTCTTTTGTGTTTGGACTCA    | 2300               |          |            |
| Db 126                | TGGAACCTGTCGTCTAATGACATATAGCATTTTCTTTTGTGTTTGGACTCA    | 67                 |          |            |
| Qy 2301               | TATTTGACATGAAGCAGATTGTTTATTAATAAATCATACATT             | 2345               |          |            |
| Db 66                 | TATTTGACATGAAGCAGATTGTTTATTAATAAATAATCATACATT          | 22                 |          |            |

```

RESULT 11
BG109895          683 bp      mRNA      linear      EST 30-JAN-2001
LOCUS             602279565F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4367128 5'
DEFINITION        mRNA sequence.
ACCESSION         BG109895
VERSION           BG109895.1 GI:12603401
KEYWORDS          EST.
SOURCE            human.
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE         1 (bases 1 to 683)
AUTHORS           NIH-MGC http://mhc.nci.nih.gov/.
TITLE             National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL           Unpublished (1999)
COMMENT           Contact: Robert Strausberg, Ph.D.
                  Email: cgapbs-remail.nih.gov
                  Tissue Procurement: ATCC
                  CDNA Library Preparation: Life Technologies, Inc.
                  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LNL at:
                  http://image.llnl.gov
                  Plate: LHAM10019 row: a column: 17
                  High quality sequence stop: 681.
                  Location/Qualifiers
                    1..683
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4367128"
                     /clone_lib="NIH_MGC_86"
                     /tissue_type="osteosarcoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bone; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.533 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC library."
BASE COUNT        232 a      125 c      145 g      180 t      1 others
ORIGIN
Query Match      27.5%; Score 645; DB 12; Length 683;
Best Local Similarity 99.8%; Pred. No. 9,4e-132;
Matches 656; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1681 ACTATTATTAACCTCTTAATATGAAACATCACTGCTTCACTGAGAACTGTATGCA 1740
DB 1 ACTATTATTAACCTCTTAATATGAAACATCACTGCTTCACTGAGAACTGTATGCA 60

QY 1741 GTGGGAAAATTAGCTGATTCATGATGATGCAATTTCTTGAATGCCACATGGAATGACACC 1800
DB 61 GTGGGAAAATTAGCTGATTCATGATGATGCAATTTCTTGAATGCCACATGGAATGACACC 120

QY 1801 CAAGTAGGGGAACGAGACTCAAGCTTTTCAGAGAGAGAAAACAAAGAGTAGCAATTGCA 1860
DB 121 CAAGTAGGGGAACGAGACTCAAGCTTTTCAGAGAGAGAAAACAAAGAGTAGCAATTGCA 180

QY 1861 AGAGCCATTTTGAAGAGACCCCCAGCTCATCTCTATGATGAAGACTTCACTGTTAGAT 1920
DB 181 AGAGCCATTTTGAAGAGACCCCCAGCTCATCTCTATGATGAAGACTTCACTGTTAGAT 240

QY 1921 TCGATTACTGAAGACACTTTCTTGTTGCCATGAAGAGTGTGTAACACAGAACTTCT 1980
DB 241 TCGATTACTGAAGACACTTTCTTGTTGCCATGAAGAGTGTGTAACACAGAACTTCT 300

QY 1981 ATTTCATTTGCACACAGATTGTCAACAGTGTGTGATGAGATGAATCATTTGCTTGAT 2040
DB 301 ATTTCATTTGCACACAGATTGTCAACAGTGTGTGATGAGATGAATCATTTGCTTGAT 360

QY 2041 CAGGTAAGTAGCCGAACGTGTATCCCACTGTTGCTTCAACCTCATAGTATC 2100

```

```

DB 361 CAGGTAAGTAGCCGAACGTGTATCCCACTGTTGCTTCAACCTCATAGTATC 420

QY 2101 TATTCAGAAATGTGCATATACAGAGACCGGTGTGAGAACCATGATTAACCCCAATGG 2160
DB 421 TATTCAGAAATGTGCATATACAGAGACCGGTGTGAGAACCATGATTAACCCCAATGG 480

QY 2161 GAACCAAGAAAGAAATATATTCCTCAAGAGAGAGAAAGAAAGAACTTCAAGAAATTT 2220
DB 481 GAACCAAGAAAGAAATATATTCCTCAAGAGAGAGAAAGAAAGAACTTCAAGAAATTT 540

QY 2221 GTCAATAGTGTGAAGAGCTGTGGAACCTGTCGTCCTAATCATATGACATTTTCTT 2279
DB 541 GTCAATAGTGTGAAGAGCTGTGGAACCTGTCGTCCTAATCATATGACATTTTCTT 600

QY 2280 TTTTGTGTTTGTGACATATTTGCACTGAAGAGAAATGTTTATTAATTAATTAATTA 2336
DB 601 TTTTGTGTTTGTGACATATTTGCACTGAAGAGAAATGTTTATTAATTAATTAATTA 657

RESULT 12
BG722908          722 bp      mRNA      linear      EST 08-MAY-2001
LOCUS             602659410F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827578 5'
DEFINITION        mRNA sequence.
ACCESSION         BG722908
VERSION           BG722908.1 GI:14002095
KEYWORDS          EST.
SOURCE            human.
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE         1 (bases 1 to 722)
AUTHORS           NIH-MGC http://mhc.nci.nih.gov/.
TITLE             National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL           Unpublished (1999)
COMMENT           Contact: Robert Strausberg, Ph.D.
                  Email: cgapbs-remail.nih.gov
                  Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
                  CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
                  Toshiyuki and Piero Carninci (RIKEN)
                  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LNL at:
                  http://image.llnl.gov
                  Plate: LHAM0744 row: c column: 03
                  High quality sequence stop: 710.
                  Location/Qualifiers
                    1..722
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4827578"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescript (modified
                     pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (9csgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
                     size-selected for average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation). Library
                     constructed by M. Brownstein (NIH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC library."
BASE COUNT        191 a      167 c      191 g      173 t
ORIGIN
Query Match      27.0%; Score 633.4; DB 12; Length 722;
Best Local Similarity 99.0%; Pred. No. 3,4e-129;
Matches 670; Conservative 0; Mismatches 1; Indels 6; Gaps 3;

QY 1 ATGGCGCTGCTCGGAGTCAATTTCTTGCGCTGGCGCGCGCGGCTGCTTTGGAAG 60
DB 44 ATGGCGCTGCTCGGAGTCAATTTCTTGCGCTGGCGCGCGCGGCTGCTTTGGAAG 103

```

QY 61 CGCGGCACTCCGCGATTCTGATCCGGCTTTAGTCTCTGTTAGCGGCTCAGGTCGCGAG 120  
Db 104 CGCGGCACTCCGCGATTCTGATCCGGCTTTAGTCTCTGTTAGCGGCTCAGGTCGCGAG 163  
QY 121 TGGAGGCGACATCACTCGGCGCTTGGGAAACCGCTGAGCTAC---CAGATTCCAGAG 177  
Db 164 TGGAGGCGACATCACTCGGCGCTTGGGAAACCGCTGAGCTACCGAGCAATTCAGAG 223  
QY 178 TCATTAATAAAGTATCAGTCAGAGATTTGGGAAAGGCAATTCAGAGCACTTCTAGAT 237  
Db 224 TCATTAATAAAGTATCAGTCAGAGATTTGGGAAAGGCAATTCAGAGCACTTCTAGAT 283  
QY 238 GCTGCAAAAGCTCTTCAGAGTATGCCACTGATAGAAAAGAGACATGTTGGCATGTCAT 297  
Db 284 GCTGCAAAAGCTCTTCAGAGTATGCCACTGATAGAAAAGAGACATGTTGGCATGTCAT 343  
QY 298 GCAGAGAGAGAGACTCCACAGAGACCCAAAGAGGTTAAAGATGTTGATACCTCGGAAA 357  
Db 344 GCAGAGAGAGAGACTCCACAGAGACCCAAAGAGGTTAAAGATGTTGATACCTCGGAAA 403  
QY 358 ATCATTAAGAAGATGCTTTCTTATGTGTGGCCCAAGACAGGCGAGATCTACGAGCTAGA 417  
Db 404 ATCATTAAGAAGATGCTTTCTTATGTGTGGCCCAAGACAGGCGAGATCTACGAGCTAGA 463  
QY 418 GTTGCCATTTGCTGGGATTTTGGGTGGTCAAAAGGCATGAATATTGTGTTCCCTTC 477  
Db 464 GTTGCCATTTGCTGGGATTTTGGGTGGTCAAAAGGCATGAATATTGTGTTCCCTTC 521  
QY 478 ATGTTTAAATATGCTGTAGAGAGGCTCAACAGATGTTCGGGAAACATGCTGAACCTAGT 537  
Db 522 ATGTTTAAATATGCTGTAGAGAGGCTCAACAGATGTTCGGGAAACATGCTGAACCTAGT 581  
QY 538 GATGACCAATAATAGTTGCAACCATGAGCAAGCAAGTTCATGTTGCTATGCTATCA 597  
Db 582 GATGACCAATAATAGTTGCAACCATGAGCAAGCAAGTTCATGTTGCTATGCTATCA 641  
QY 598 AGAGCTGAGC-TGCTTTTAAAGAAAGTTGCAATGCAATGTTGGCAAGGTAGCCCA 656  
Db 642 AGAGCTGAGCTGCTTTTAAAGAAAGTTGCAATGCAATGTTGGCAAGGTAGCCCA 701  
QY 657 GAATTCATCCGAGAA 673  
Db 702 GAATTCATCCGAGAA 718

RESULT 13  
BF310056 817 bp mRNA linear EST 21-NOV-2000  
LOCUS  
DEFINITION 601892556F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4138236 5',  
mRNA sequence.  
ACCESSION BF310056  
VERSION BF310056.1 GI:11257555  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 817)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph. D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
plate: LLCM046 row: p column: 13  
High quality sequence stop: 644.  
Location/Qualifiers

source 1. 817  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4138236"  
/clone\_1ib="NIH MGC 17"  
/cistruc\_type="rhabdomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pORF7; Site 1: EcoRI;  
Site 2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 207 a 192 c 227 g 191 t  
ORIGIN  
Query Match 26.6%; Score 623.2; DB 12; Length 817;  
Best Local Similarity 96.8%; Pred. No. 5.9e-127;  
Matches 722; Conservative 0; Mismatches 13; Indels 11; Gaps 8;  
QY 1 ATGGCGCTGCTCCGCGATTCCTTGGCGGCGCGCGCGCGCTCTTTCGAAAAG 60  
Db 3 ATGGCGCTGCTCCGCGATTCCTTGGCGGCGCGCGCGCGCTCTTTCGAAAAG 62  
QY 61 CGCGGCACTCCGCGATTCCTTGGCGGCGCGCGCGCGCTCTTTCGAAAAG 120  
Db 63 CGCGGCACTCCGCGATTCCTTGGCGGCGCGCGCGCGCTCTTTCGAAAAG 122  
QY 121 TGGAGGCGACATCACTCGGCGCTTGGGAAACCGCTGAGCTTAC---CAGATTCCAGAG 177  
Db 123 TGGAGGCGACATCACTCGGCGCTTGGGAAACCGCTGAGCTTACCGAGATTCAGAG 182  
QY 178 TCATTAATAAAGTATCAGTCAGAGATTTGGGAAAGGCAATTCAGAGCACTTCTAGAT 237  
Db 183 TCATTAATAAAGTATCAGTCAGAGATTTGGGAAAGGCAATTCAGAGCACTTCTAGAT 241  
QY 238 GCTGCAAAAGGCTCTCCAGAGTATGAGCAATGAAAAGAGACATGTTGGCATGTCAT 297  
Db 242 GCTGCAAAAGGCTCTCCAGAGTATGAGCAATGAAAAGAGACATGTTGGCATGTCAT 300  
QY 298 GCAGAGAGAGAGACTCCACAGAGACCCAAAGAGGTTAAAGATGTTGATCTCGGAAA 357  
Db 301 GCAGAGAGAGAGACTCCACAGAGACCCAAAGAGGTTAAAGATG-TGATATCTCGGAAA 359  
QY 358 ATCATTAAGAAGATGCTTTCTTATGTGTGGCCCAAGACAGGCGAGATCTACGAGCTAGA 417  
Db 360 ATCATTAAGAAGATGCTTTCTTATGTGTGGCCCAAGACAGGCGAGATCTACGAGCTAGA 419  
QY 418 GTTGCCATTTGCTGGGATTTTGGGTGGTCAAAAGGCATGAATATTGTGTTCCCTTC 477  
Db 420 GTTGCCATTTGCTGGGATTTTGGGTGGTCAAAAGGCATGAATATTGTGTTCCCTTC 479  
QY 478 ATGTTTAAATATGCTGTAGAGAGGCTCAACAGATGTCGGGAAACATGCTGAACCTGAGT 537  
Db 480 ATGTTTAAATATGCTGTAGAGAGGCTCAACAGATGTCGGGAAACATGCTGAACCTGAGT 539  
QY 538 GATGACCAATAATAGTTGCAACCATGAGCAAGCAAGTTCATGTTGCTATGCTATCA 597  
Db 540 GATGACCAATAATAGTTGCAACCATGAGCAAGCAAGTTCATGTTGCTATGCTATCA 598  
QY 598 AGAGCTGAGAGCTCTTTTAAAGAAAGTTGCAATGCAATGTTGGCAAGGTAGCCCAAG 657  
Db 599 AGAGCTGAGAGCTCTTTTAAAGAAAGTTGCAATGCAATGTTGGCAAGGTAGCCCAAG 657  
QY 658 AATTCAATCCGAAGATAGCAAAAATGCTTTC-TCCATCTTCACAA--CCTGAGATCTG 714  
Db 658 AATTCAATCCGAAGATAGCAAAAATGCTTTC-TCCATCTTCACAAACCTGATCTCTG 717  
QY 715 GGTTCACCTTACAG 740  
Db 718 GGTTCACCTTACAG 743

RESULT 14  
LOCUS BE618331 758 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601462724P1 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:386633 5',  
mRNA sequence.  
ACCESSION BE618331  
VERSION BE618331.1 GI:9889269  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 758)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LLM9610 row: 0 column: 06  
High quality sequence stop: 648.  
Location/Qualifiers  
source 1..758  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:386633"  
/clone\_lib="NIH\_MGC\_67"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-Sport6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: oligo  
Average insert size 1.75 kb. Library constructed by Life  
Technologies."  
BASE COUNT 214 a 165 c 195 g 184 t  
ORIGIN  
Query Match 26.2%; Score 614; DB 10; Length 758;  
Best Local Similarity 94.4%; Pred. No. 6.3e-125;  
Matches 714; Conservative 0; Mismatches 30; Indels 12; Gaps 7;  
QY 74 CGATTCTGATCCGGCTTGTAGTCTCTGTAGCGGCTCAAGTCCGAGTGGAGCCACATC 133  
Db 1 CGATTCTGATCCGGCTTGTAGTCTCTGTAGCGGCTCAAGTCCGAGTGGAGCCACATC 60  
QY 134 AACTGGCGCCCTTGGGAAACCGCTCGAGCTTAC----CAGATTCCAGAGTCATTAAAAAGT 189  
Db 61 AACTGGCGCCCTTGGGAAACCGCTCGAGCTTACCGAGCAGATTCAGAGTCATTAAAAAGT 120  
QY 190 ATCAGATGGCAAGATTTGGGAAAGGCAATTAGACAGTCTTGTAGTCTGCAAAAGCT 249  
Db 121 ATCAGATGGCAAGATTTGGGAAAGGCAATTAGACAGTCTTGTAGTCTGCAAAAGCT 180  
QY 250 CTCGAGTATGGCCACTGATAGAAAAAGAGCATGTTGGCATGTGCATGACAGAGAGGA 309  
Db 181 CTCGAGTATGGCCACTGATAGAAAAAGAGCATGTTGGCATGTGCATGACAGAGAGGA 240  
QY 310 CTCGACACAGACCCCAAAAGAGGTTAAAGATGTTGATCTCGAAAAATCTAAAAAGA 369  
Db 241 CTCGACACAGACCCCAAAAGAGGTTAAAGATGTTGATCTCGAAAAATCTAAAAAGA 300  
QY 370 ATGCTTCTTATATGTGGGCGCAAAAGACGCGCAGATCTACGAGTAAAGTTGCCATTTCG 429  
Db 301 ATGCTTCTTATATGTGGGCGCAAAAGACGCGCAGATCTACGAGTAAAGTTGCCATTTCG 360  
QY 430 CTGGATTCTTTGGTGTGTGCAAAAGGCATGAATATTGTGTTCCCTTCATGTTTAAATAT 489

RESULT 15  
LOCUS A1769898/c 627 bp mRNA linear EST 21-DEC-1999  
DEFINITION w330c01.x1 NCI CGAP Kid12 Homo sapiens cDNA clone IMAGE:2404320 3',  
similar to SW:ABC7 HUMAN 075027 ATP-BINDING CASSETTE TRANSPORTER 7  
PRECURSOR ;, mRNA sequence.  
ACCESSION A1769898  
VERSION A1769898.1 GI:5236407  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 627)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert length: 556 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 473.  
Location/Qualifiers  
source 1..627  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2404320"  
/clone\_lib="NCI-CGAP\_Kid12"  
/tissue\_type="2 pooled tumors (clear cell type)"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP Kids was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subclative  
hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library  
(cloneIDs 1325912-1325912, 1471368-1472903 and  
1492104-1493255). Subtraction by Bento Soares and M.  
Fátima Bonaldo.

BASE COUNT 164 a 140 c 111 g 212 t  
ORIGIN

Query Match 26.0%; Score 610.4; DB 9; Length 627;  
Best Local Similarity 99.0%; Pred. No. 4e-124;  
Matches 614; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1726 GAGGAGGTGATGCAAGTGGCAAAATTAGCTGACTTCTGATGCAATTTCTGAAATGCCA 1785  
DB 627 GAGGAGGTGATGCAAGTGGCAAAATTAGCTGACTTCTGATGCAATTTCTGAAATGCCA 568  
QY 1786 CATGATATGACACCCCAAGTGGGGAACGAGACTCAAGCTTTTCAGAGAGAGAAAGCAA 1845  
DB 567 CATGATATGACACCCCAAGTGGGGAACGAGACTCAAGCTTTTCAGAGAGAGAAAGCAA 508  
QY 1846 AGAGTAGCAATTCAGAGAGCAATTTGAAAGACCCCACTCATATCTATGATGAAGCT 1905  
DB 507 AGAGTAGCAATTCAGAGAGCAATTTGAAAGACCCCACTCATATCTATGATGAAGCT 448  
QY 1906 ACTTCATGCTTGAATTCGATTAAGAGACTATTTCTGTCCTATGAAGATGTGTC 1965  
DB 447 ACTTCATGCTTGAATTCGATTAAGAGACTATTTCTGTCCTATGAAGATGTGTC 388  
QY 1966 AAACACAGAACTTCTATTTTATTCACACAGATTTGCAACAGTGTGATGCAATGAA 2025  
DB 387 AAACACAGAACTTCTATTTTATTCACACAGATTTGCAACAGTGTGATGCAATGAA 328  
QY 2026 ATCATTTGCTTGAATTCAGGTTAGGTAGCCGAGCTGTTACCAACATGTTGCTTCT 2085  
DB 327 ATCATTTGCTTGAATTCAGGTTAGGTAGCCGAGCTGTTACCAACATGTTGCTTCT 268  
QY 2086 AACCTCATATATCTATTTTCAAAATGTCATACAGAGACGCTGTGCAAGACCAT 2145  
DB 267 AACCTCATATATCTATTTTCAAAATGTCATACAGAGACGCTGTGCAAGACCAT 208  
QY 2146 GATAACCCCAATGGGAAGCAAGAAATAATATCCAAAGAGAGAGAAAGAAA 2205  
DB 207 GATAACCCCAATGGGAAGCAAGAAATAATATCCAAAGAGAGAGAAAGAAA 148  
QY 2206 CTACAGAGAGAAATGTCATATAGTGTGAAAGCTGTGAAACTGTCTGCTAAGTACA 2265  
DB 147 CTACAGAGAGAAATGTCATATAGTGTGAAAGCTGTGAAACTGTCTGCTAAGTACA 88  
QY 2266 TAAGACATTTTCTTTTGTGTTTGTGACTACATATTTGCACTGAAGCAAAATGTTT 2325  
DB 87 TAAGACATTTTCTTTTGTGTTTGTGACTACATATTTGCACTGAAGCAAAATGTTT 28  
QY 2326 TATTAAAAAATCATACATT 2345  
DB 27 TATTAAAAAATCATACATT 8

Search completed: February 15, 2003, 05:49:04  
Job time : 3269 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 15, 2003, 04:58:20 ; Search time 88 Seconds

(without alignments)  
10981.375 Million cell updates/sec

Title: AF133659  
Perfect score: 4254  
Sequence: 1 ATGGCCCTGCTCGCATGCA.....TATTAAAAATCATACATT 2345

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p.model -DEV=xih  
-O=cpm2.1/USPTO.spool/AF133659/runat.10022003.155441.29523/app.query.fasta.1.2503  
-DB=SPTRMBL\_21 -QFMT=fastan -SUFFIX=epc -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=AF133659 @CGN 1.128 @runat.10022003.155441.29523 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL\_21:\*  
1: sp archaea:\*  
2: sp bacteria:\*  
3: sp fungi:\*  
4: sp human:\*  
5: sp invertebrate:\*  
6: sp mammal:\*  
7: sp mhc:\*  
8: sp organelle:\*  
9: sp phage:\*  
10: sp plant:\*  
11: sp rodent:\*  
12: sp virus:\*  
13: sp vertebrate:\*  
14: sp unclassified:\*  
15: sp tvirus:\*  
16: sp bacteriaph:\*  
17: sp archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

| Result No. | Score | Query Match | Length | DB ID    | Description       |
|------------|-------|-------------|--------|----------|-------------------|
| 1          | 1852  | 43.5        | 606    | 5 Q9W0C5 | Q9W0C5 drosophila |

|    |        |      |      |    |         |                     |
|----|--------|------|------|----|---------|---------------------|
| 2  | 1727.5 | 40.6 | 728  | 10 | Q91WML  | Q91WML arabidopsis  |
| 3  | 1715.5 | 40.3 | 728  | 10 | Q91F78  | Q91F78 arabidopsis  |
| 4  | 1657.5 | 39.0 | 677  | 10 | Q9M0G8  | Q9M0G8 arabidopsis  |
| 5  | 1657.5 | 39.0 | 678  | 10 | Q9FUT3  | Q9FUT3 arabidopsis  |
| 6  | 1620.5 | 38.1 | 680  | 10 | Q9M0G9  | Q9M0G9 arabidopsis  |
| 7  | 1599   | 37.6 | 697  | 5  | Q8T9W2  | Q8T9W2 dicystoseli  |
| 8  | 1577.5 | 37.1 | 936  | 10 | Q9S8N5  | Q9S8N5 oryza sativ  |
| 9  | 1454.5 | 34.2 | 631  | 16 | Q82D10  | Q82D10 yersinia pe  |
| 10 | 1446.5 | 34.0 | 609  | 16 | Q9ZDM0  | Q9ZDM0 rickettsia   |
| 11 | 1400   | 32.9 | 628  | 16 | Q8YF88  | Q8YF88 bruceella me |
| 12 | 1399   | 32.9 | 592  | 16 | Q88258  | Q88258 ralsstonia s |
| 13 | 1393.5 | 32.8 | 627  | 16 | Q984W3  | Q984W3 rhizobium l  |
| 14 | 1375   | 32.3 | 629  | 16 | Q8UGH3  | Q8UGH3 agrobacteri  |
| 15 | 1369   | 32.2 | 627  | 16 | Q92R12  | Q92R12 rhizobium m  |
| 16 | 1247   | 29.3 | 836  | 11 | Q70595  | Q70595 rattus norv  |
| 17 | 1241.5 | 29.2 | 842  | 11 | Q9DC29  | Q9DC29 mus musculu  |
| 18 | 1240.5 | 29.2 | 766  | 4  | Q96ME8  | Q96ME8 homo sapien  |
| 19 | 1239.5 | 29.1 | 896  | 4  | Q9HAC7  | Q9HAC7 homo sapien  |
| 20 | 1183   | 27.8 | 866  | 5  | Q9VFE20 | Q9VFE20 drosophila  |
| 21 | 1042   | 24.5 | 643  | 16 | Q9A8N9  | Q9A8N9 caulobacter  |
| 22 | 923.5  | 21.7 | 1025 | 5  | Q25693  | Q25693 plasmodium   |
| 23 | 915    | 21.5 | 704  | 5  | Q9XUJ1  | Q9XUJ1 caenorhabdi  |
| 24 | 908    | 21.3 | 947  | 5  | Q08667  | Q08667 plasmodium   |
| 25 | 799.5  | 18.8 | 320  | 5  | Q955T0  | Q955T0 caenorhabdi  |
| 26 | 725    | 17.0 | 642  | 17 | Q9HSQ7  | Q9HSQ7 halobacteri  |
| 27 | 709    | 16.7 | 581  | 16 | Q9KDT6  | Q9KDT6 bacillus ha  |
| 28 | 708    | 16.6 | 597  | 16 | Q8YV35  | Q8YV35 anabaena sp  |
| 29 | 706.5  | 16.6 | 593  | 16 | P73239  | P73239 synecocyst   |
| 30 | 700    | 16.5 | 832  | 16 | Q9A2Y6  | Q9A2Y6 caulobacter  |
| 31 | 699.5  | 16.4 | 578  | 16 | Q99T13  | Q99T13 stephylcococ |
| 32 | 693.5  | 16.3 | 725  | 16 | Q8XT23  | Q8XT23 ralsstonia s |
| 33 | 690.5  | 16.2 | 1243 | 16 | Q9RDX5  | Q9RDX5 streptomyce  |
| 34 | 687    | 16.1 | 585  | 16 | Q31707  | Q31707 bacillus su  |
| 35 | 682.5  | 16.0 | 735  | 2  | Q9F7V7  | Q9F7V7 rhizobium l  |
| 36 | 678.5  | 15.9 | 851  | 16 | Q9RYG8  | Q9RYG8 thermococcus |
| 37 | 676.5  | 15.9 | 594  | 2  | P74940  | P74940 thermotoga   |
| 38 | 675.5  | 15.9 | 851  | 17 | Q972N4  | Q972N4 sulfolobus   |
| 39 | 673.5  | 15.8 | 600  | 16 | Q8U6S8  | Q8U6S8 agrobacteri  |
| 40 | 673.5  | 15.8 | 863  | 17 | Q97UJ9  | Q97UJ9 sulfolobus   |
| 41 | 673    | 15.8 | 615  | 16 | Q9CLD7  | Q9CLD7 pasteurella  |
| 42 | 672.5  | 15.8 | 546  | 16 | Q8Y247  | Q8Y247 anabaena sp  |
| 43 | 672.5  | 15.8 | 707  | 2  | Q47258  | Q47258 escherichia  |
| 44 | 672    | 15.8 | 708  | 2  | Q93FG4  | Q93FG4 manheimia    |
| 45 | 672    | 15.8 | 708  | 2  | Q93313  | Q93313 manheimia    |

## ALIGNMENTS

## RESULT 1

| ID | Q9W0C5  | PRELIMINARY; | PRT; | 606 AA. |
|----|---|--------------|------|---------|
| AC | Q9W0C5;   |              |      |         |
| DT | 01-MAY-2000 (TREMBLrel. 13, Created)                                  |              |      |         |
| DT | 01-MAY-2000 (TREMBLrel. 13, Last sequence update)                     |              |      |         |
| DT | 01-JUN-2002 (TREMBLrel. 21, Last annotation update)                   |              |      |         |
| DE | CG7955 protein (GH20617p).  |              |      |         |
| DE | CG7955  |              |      |         |
| GN | Drosophila melanogaster (Fruit fly).                                  |              |      |         |
| OS | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;         |              |      |         |
| OC | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; |              |      |         |
| OC | Ephydroidea; Drosophilidae; Drosophila.                               |              |      |         |
| OX | NCBI_TaxID=7227;  |              |      |         |
| OX | [1]   |              |      |         |
| RP | SEQUENCE FROM N.A.  |              |      |         |
| RC | STRATIN=BERKELEY;   |              |      |         |
| RX | MEDLINE=20196006; PubMed=10731132;                                    |              |      |         |
| RA | Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,       |              |      |         |
| RA | Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,     |              |      |         |
| RA | George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,   |              |      |         |
| RA | Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,         |              |      |         |
| RA | Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  |              |      |         |
| RA | Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., |              |      |         |
| RA | Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D., |              |      |         |



RA Ballew R.M., Baau A., Baxendale J., Bayraktaroglu I., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Bokora D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Butris K.C., Buam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwan C.,  
 RA Jalali M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kalp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskeim D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kimos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Mu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhao Q., Zheng L.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-BERKELEY.  
 RC  
 RA Stapleton M., Brokstein P., Hong L., Asgharani A., Carlson J.,  
 RA Champe C., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;  
 RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL; AE003472; AAF47525.1; -;  
 DR EMBL; AY01556; AAK92980.1; -;  
 DR FLYBase; FBgn0035244; CG7955.  
 DR InterPro; IPR0013593; AAA\_Arpase.  
 DR InterPro; IPR001140; ABCtransporter.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00664; ABC\_membrane; 1.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR KW ATP-binding; Transport.  
 SQ SEQUENCE 606 AA; 66104 MW; 76C8050A1B62327 CRC64;

Alignment Scores:  
 Pred. No.: 3,31e-120 Length: 606  
 Score: 1852.00 Matches: 360  
 Percent Similarity: 76.92% Conservative: 110  
 Best Local Similarity: 58.92% Mismatches: 131  
 Query Match: 43.54% Indels: 10  
 DB: 5 Gaps: 3

AF133659 (1-2345) x Q9W0C5 (1-606)  
 QY 358 ATCATATAAGCATCTTCTTATGTGTGGCCCAAGACGCGCATCTACGACTGAGA 417  
 Db 1 MetLeuAArgAlaMetMeChalTArTleTprProLysGluAspProLeuValArgLysArg 20

QY 418 GTTGCATTTTCCCTGGGATTTTGGGTGGTGCAGAAAGGCATATAATTTGGTTCCCTTC 477  
 Db 21 ValGlyIleSerLeuGlyLeuLeuAlaGlySerLysLeuLeuThValCysValProPhe 40

QY 478 ATGTTAAATATGCTGTAGACAGCCTCAGCAGATGTCGGGAAAATGCTGAACCTGAGT 537  
 Db 41 LeuHeuLysGlyValValAlaAspThMet-----ThrThrLeuAsnMetLeuAsp 55  
 QY 538 GATGCACCAATATACAGTTTGCACACATGGCAACAGCACTTCTGATGGCTATGCTATCA 597  
 Db 56 ThrIleProAlaPalaValLeuSerAlaIleThrIleLeuMetLeuGlyTyrGlyIleAla 75  
 QY 598 AGACGTGGAGCTGCTTTTAAAGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 657  
 Db 76 ArgAlaSerAlaAlaGlyPheAsnGluLeuArgAsnAlaValAlaPheAlaLysValAlaHis 95  
 QY 658 AATTCATCCGAGAGATAGCCCAAAATGCTCTTCCATCTTCAACCTGAGATCGAGT 717  
 Db 96 HisSerIleArgLysIleAlaSerAsnValPheLeuHisLeuHisAsnLeuAspLeuAla 115  
 QY 718 TTTCACCTGACAGACAGACGCGGAGCTTTATCTAAGCTATTGACAGAGAACAGAGGCT 777  
 Db 116 PheHisLeuAsnLysGlnThGlyAlaLeuSerLysThrIleAspArgGlySerArgGly 135  
 QY 778 ATCAGTTTTCCTGAGGCTTGTGATTAATCTTCCATCATGTTGAAGTATG 837  
 Db 136 IleAsnPheValLeuSerAlaMetValPheAsnIleValProThrIlePheGluLeuAla 155  
 QY 838 CTGTGACAGTGTGTTTGTATTAATTAATGCGGTGCGCAGTTTGTGTAACCTTGA 897  
 Db 156 LeuValSerSerIleLeuGlyValLysCysGlyLeuAlaPheAlaGlyValSerMetGly 175  
 QY 898 AACTGTGTATATACACAGCATTTACAGCTTGCAGCTCACCGGTGGAAGACTGATTAGA 957  
 Db 176 CysValGlyIleTyrAlaAlaTyrThrLeuSerValThrGlnTrpArgThGlnPheArg 195  
 QY 958 ATAGAAATGAACAAGAGATATGATGACAGTATGCTGCTATGATGACTGCTGAT 1017  
 Db 196 ValPheMetAsnGlnAlaGluAsnGluAlaGlyAsnLysAlaValAspSerLeuIleAsn 215  
 QY 1018 TATGAACCTGGAAGTATTTAATTAATGAAGATATGAAGACACAGATATGATGATTT 1077  
 Db 216 TyrGlnThrValLysTyrPheAsnAsnGlnLysTyrGlnAlaGlyCysTyrAsnGluVal 235  
 QY 1078 TTGAAGACGTATGAGACTGCTTCACTTGAAGAAGTACCTTACTGCTGCTATGCTGA 1137  
 Db 236 LeuLysLysTyrGluAlaAlaSerLeuLysThrSerSerSerLeuAlaLeuLeuAsnPhe 255  
 QY 1138 GGTCAAGTCTATTTTTCAGTGTGCTTTTAAACAGCTTAATGAGTGTGCGCAGCAGGA 1197  
 Db 256 GlyLysAsnAlaIlePheSerSerAlaLeuSerLeuIleMetValLeuAlaIleLysGlu 275  
 QY 1198 ATTGTGACAGGTACCTTACTGTGTGAGATCTTAATGTGAATGAGCTGCTTTTCAG 1257  
 Db 276 IleAlaGlnGlyAsnMetThrValGlyAspLeuValMetValAsnAlaLeuLeuPheGln 295  
 QY 1258 CTTTCATTAACCTTGAACCTTCTGGGAATCTGTATATAGAGACTAGACAAAGCACTATA 1317  
 Db 296 LeuSerIleProLeuGlyPheLeuGlySerValTyrArgGluValArgGlnAlaLeuLeu 315  
 QY 1318 GATATGAACACTGTTTACTCTCTCAAGTATGACCCCAATTAATGAACAAGTATG 1377  
 Db 316 AspPheArgAlaMetPheThrLeuMetAsnAlaAspSerSerIleGlnThrAlaAlaAsn 335  
 QY 1378 GCATCTCCCTTGCATGACACACAGACAGACTACGCTGCGCTTGTATATGATGATTT 1437  
 Db 336 AlaGlnProLeuPheValAspThrThrAsnSerSerIleGluPheArgAsnValSerPhe 355  
 QY 1438 GAATACATTGAGGCCAGAAAGTCTTAGTGGAATATCTTGAAGTCCCTGCGAGGAAG 1497  
 Db 356 GluTyrGluProGlyLysProIlePheArgAspLeuSerPheThrIleProAlaGlyLys 375  
 QY 1498 AAGTGGCATTTGAGAGAGTATGAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1557  
 Db 376 AsnValAlaIleValGlyGlySerGlySerGlySerGlySerSerMetValArgLeuLeuPhe 395  
 QY 1558 CGCTTCTATGAGCTCAAAAGGATGATTTATCTGCTGATCAAAATATACAGATGTG 1617



```

Db 161 oPhenuleuPheValAlaAspTrpLeuAlaSerAlaThrGlyThrGlyAlaSerLe 181
Qy 534 GAGTATGACCAAAATATACAGTTGCACCATGGCAACA-----GCAGT 575
Db 181 uThrTrpPheAlaAlaAlaAsnProThrLeuThrValPheAlaThrProAlaAlaVa 201
Qy 576 TCTGATGGCTATGGTGTATCAAGAGCTGAGCTGCTTTTATTAAGAAAGTTGAAATGC 635
Db 201 lLeuIleGlyTrgIlyIleAlaArgThrGlySerSerAlaPheAsnGluLeuArgThrAl 221
Qy 636 AGTATTTGGCAAGGTAGCCCAAGAAATTCATCCGAGAAATAGCCAAAATGCTTTCTCCA 695
Db 221 aValPheSerIysValAlaLeuArgThrIleArgSerValSerArgIysValPheSerHi 241
Qy 696 TCTTCACAACTGGATCTGGGTTTTCACCTGAGCAGACAGAGGAGCTTATCTAAGCC 755
Db 241 sLeuHisAspLeuAspLeuArgIlyrHisLeuSerArgIlyrGlyLeuAsnArgI 261
Qy 756 TATTGACAGAGGAACAAGGGGTATCAGTTTGTCTGAGTGGCTGATTTAATCTCTCT 815
Db 261 eileAspArgIlySerAlaGlyAlaIleAsnPheIleuSerAlaMetValPheAsnVala 281
Qy 816 TCCCATCATGTTGAAGATGCTGTGTCAGTGGTGTGTTGATTAACAATGCGGTGCCA 875
Db 281 lProThrIleuGluIleSerMetValSerGlyIleLeuAlaTyrIysPheGlyAlaAl 301
Qy 876 GTTTCCTTGGTAACCTTGGAAACCTTGATACATACACAGATTCACAGTTGCAGTAC 935
Db 301 aPheAlaTrpIleThrSerLeuSerValGlySerIlyrIleValPheThrIleAlaValTh 321
Qy 936 ACGGAGGAACTAGATTAGAAATGAATGAACAAGACAGATATGATGCGATATGC 995
Db 321 rGlnTrpArgThrIlySerPheArgIysAlaMetAsnIysAlaAspAsnAlaSerThrAr 341
Qy 996 TGTATAGACTCATCTGCAATATATGAACCTGTGAAGTATTTAAATGAAGAATATGA 1055
Db 341 gAlaIleAspSerLeuIleAsnTyrGluThrValIlyrPheAsnAsnGluGlyTyrG 361
Qy 1056 AGCAGAGAGATATGATGATTTTGAAGACGATAGACTGCTTATGAAAAGTACCTC 1115
Db 361 uAlaGluIlyrIlyrAspGlnPheLeuIlySerIlyrGluAspAlaIleuGlnThrGlnAr 381
Qy 1116 TACTGTGGCTATGCTGAACTTTGTCAAAGTCTATTTTTCAGTGCCTGTTAACAGCTAT 1175
Db 381 gSerIleuAlaPheLeuAsnPheGlyGlnSerIleIlePheSerThrAlaLeuSerThrAl 401
Qy 1176 AATGTGCTCGCCAGTACAGGAATTTGGCAGACCTTACTGTTGAGATCTAGTAAT 1235
Db 401 aMetValLeuCysSerGlnGlyIleMetAsnGlyIlyMetThrValGlyAspLeuValMe 421
Qy 1236 GGTGAATGAGCTGCTTTTTCAGCTTTTCATTAACCTTCGAACTTCTGGGAACGTATATAG 1295
Db 421 tValAsnIlyLeuLeuPheGlnLeuSerLeuProLeuAsnPheLeuGlySerValTyrAr 441
Qy 1296 AGAGACTAGACAGACTCATATGATATGAACCTTGTTTACTTACTCAAGGTAGACAC 1355
Db 441 gGluThrIleGlnSerLeuValAspMetIlySerMetPheGlnLeuLeuGluGlyLysE 461
Qy 1356 CCAATTAAGAACAAGTATGACATCTCCCTTCAGATCACACACAGACAGTACCGT 1415
Db 461 rAspIleThrAsnThrSerAspAlaIlyrProLeuValLeu-----LysGlyGlyAsnIl 479
Qy 1416 GGCCTTGATATATGTCATTTGAAATACATTGAGGCCCAAGAAATCTTATGGAATATC 1475
Db 479 eGluPheGluAsnValHisPheSerTyrLeuProGluArgIlyrIleLeuAspGlyIleSe 499
Qy 1476 CTTTGAAGTCCCTGAGGAAAGATGGCCATTTAGAGAGTATGGGTACGGGAAAG 1535
Db 499 rPheValValProAlaGlyLysSerValAlaIleValGlyThrSerGlySerGlyLysE 519
Qy 1536 CACATAGTGAAGCTATATATTCGCTTCATAGAGCTCAAAAGGTACATTTATCTTGC 1595

```

```

Db 519 rThrIleLeuArgMetLeuPheArgPhePheAspThrAspSerGlyAsnIleArgIleAs 539
Qy 1596 TGTCAAAATATATACAGATGAGCCCTGAAAGCTTCGAGAGGACATGGAGATGATACC 1655
Db 539 pGlyGlnAspIleLeuGluValAlaArgLeuAspSerLeuArgSerSerIleGlyValIlyPr 559
Qy 1656 TCAGATGCTGTCTCTTTCATATACATATTTATTAACAACCTCTTATATAGAAACATCAG 1715
Db 559 oGlnAspThrValLeuPheAsnAspThrIlePheHisAsnIleHisIlyrGlyArgLeuSe 579
Qy 1716 TGTTCACCTGAGGAAGTATGATGACGTGCGAAATTAAGTCTGACATTATGATGCAATTC 1775
Db 579 rAlaThrGluGluGluValValYrGluAlaAlaArgArgAlaAlaIleHisGluThrIleSe 599
Qy 1776 TCGAATGCCATGATATGATGACCCCAAGTAGGGGAACGAGACTCAAGCTTCAGAGAG 1835
Db 599 rAsnPheProAspIlyrIlyrSerThrIleValGlyGluArgGlyLeuIlySerGlyG 619
Qy 1836 AGAAAGCAAAAGACTAGCAATTCGACAGACCATTTTGAAGACCCCACTCATCTCTA 1895
Db 619 yGluIlyGlnArgValAlaIleuAlaIleuAlaArgThrPheLeuIlySerProAlaIleLeuLeuCy 639
Qy 1896 TGATGAAGCTACTTCATCCTGATGATTTGATTCATGAAAGACATTTCTTGCGCATGAA 1955
Db 639 sAspGluAlaThrSerAlaLeuAspSerThrThrGluAlaGluIleuAsnAlaLeuIy 659
Qy 1956 GGATGTGGTCAACACAGAACTTATTTTCATTCGACACAGATGTGCAACAGTGTGA 2015
Db 659 sAlaLeuAlaSerAsnArgThrSerIlePheIleAlaHisArgLeuThrThrAlaMetG 679
Qy 2016 TGCAGATGAATCATTTGCTTGGATCAGGGTAAAGTAGGCAAGCTGTATGCCATGCG 2075
Db 679 nCysAspGluIleValValLeuGluAsnGlyLysValValGluGlnGlyProHisAspG 699
Qy 2076 TTGCTTGTCTAACCTCATATGATCTTTCAGAAATGTGGCATACAGAGACCGCTGT 2135
Db 699 uLeuLeuGlyLys---SerGlyArgTyrAlaGlnLeuTrpThrGlnGlnAsnSerSerVa 718
Qy 2136 GCAGAACCATGAT 2148
Db 718 lAspMetLeuAsp 722

RESULT 3
Q9LF78 PRELIMINARY; PRT; 728 AA.
AC Q9LF78;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Mitochondrial half-ABC transporter..
GN STAL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096387; PubMed=11158531;
RA Kishnir S., Babychuk E., Storozenko S., Davey M.W., Papenbrock J.,
RA De Rycke R., Engler G., Stephan U.W., Kispal G., Lill R.,
RA Van Montagu M.;
RT "A mutation of the Mitochondrial ABC transporter Stal leads to
RT dwarfism and chlorosis in the Arabidopsis mutant sterik.";
RL Plant Cell 13:89-100(2001).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AJ272202; CAB97048.1; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001140; ABCtransportTM.
DR InterPro; IPR003439; ABC_transportr.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transportr; 1.

```

DR SMART: SM00382; AAA, 1  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 DR ATP-binding; Transport.  
 SQ SEQUENCE 728 AA; 80394 MW; 29FESAAA2F77949C CRC64;

## Alignment Scores:

| Pred. No.:             | 9.81e-111 | Length:       | 728 |
|------------------------|-----------|---------------|-----|
| Score:                 | 1715.50   | Matches:      | 364 |
| Percent Similarity:    | 65.64%    | Conservative: | 125 |
| Best Local Similarity: | 48.86%    | Mismatches:   | 190 |
| Query Match:           | 40.33%    | Indels:       | 66  |
| DB:                    | 10        | Gaps:         | 10  |

AF133659 (1-2345) x Q9LF78 (1-728)

```

QY CGCGCGCGCTCTTCGAAAAGCCGGCAGCTC----- 71
Db 3 ArgGlySerArgPheValArgAlaProGlyLeuLeuLeuCyArgValAsnLeuGlnPro 22
QY 72 CGCGATTCTGATCCGGCCTTAGTCTGTGTTAGCGGCTCAGTCCGACGTGAGGCCACA 131
Db 23 GlnProLysIleProSerPheSerTyrSerLeuArg----- 34
QY 132 TCAACTCGCGCCCTTTGGGAAC-----CGCTGAGCCTTACAGAT 170
Db 35 SerAspTyrArgLeuHisAsnGlyPheSerAsnTyrIleArgArgAsnSerIleArgThr 54
QY 171 TCCAGAGTCATTAAAGATAT----- 191
Db 55 SerProValIleAsnAlaPheLeuSerAspAsnSerProSerProSerProSerProSer 74
QY 192 -----CACATGCGCAGAGATTGGGAAAGCAATTGAGACAGATTCTT 233
Db 75 ProIleArgPheValGlnArgSerSerMetLeuAsnGlyArgLeuPheSerThr----- 92
QY 234 AGATGCTGCAAAAGCTCTCCAGATATGCCCTGATAGAAAAGAGACATTGTGCATGG 293
Db 93 -----SerThrProAsnProAspGlnThrThrIleThrIleLysGluIle----- 107
QY 294 TCATGACGAGAGAGGATCCACACAGACCCCAAGAAAGGTTAAAGATGTTGATCTCG 353
Db 108 -----LysThrThr-SerSerAspSerAspSerAlaMetAlaAspMet----- 121
QY 354 GAAATATCATAAAGCAATGCTTTCTTATGTGTGGCCCAAGACAGCCAGATCTTACGAGC 413
Db 122 -LysIleLeuArgThrLeuAlaGlyTyrLeuTyrMetArgAspAsnProGlnPheArgPn 141
QY 414 TAGAGTTGCCATTTCGCTGGGATTTTGGGTGGTGCAGAGGCCATGAATATGTGTTCC 473
Db 141 eArgValIleAlaAlaLeuGlyPheLeuValGlyAlaIleValLeuAsnValGlnValPr 161
QY 474 CTTTCAATTTAAATATATGCTAGACAGCTTCACAGATGTCGGGAAACATGCTGAACCT 533
Db 161 orPheLeuPheLeuLeuAlaValAspTyrPheAlaSerAlaThrGlyAlaSerLe 181
QY 534 GAGTATGACCAACCAATACAGTTGCACACCATGCAACAGATGCGGAAACATGCTGAACCT 575
Db 181 uThrThrPheAlaAlaThrAsnProThrLeuLeuThrValPheAlaThrProAlaAlaVa 201
QY 576 TCTGATTTGGCTATGCTGTATCAAGAGTGAAGTCTTTTAAAGAAAGTTCGAAATGC 635
Db 201 IleuIleGlyTyrGlyIleAlaArgThrGlySerSerAlaPheAsnGlnLeuArgThrAl 221
QY 636 AGTATTTGGGAGGTAGCCCAAGATTCATTCGGAAGAAATAGCCAAAATGCTTTCTTCCA 695
Db 221 AlaAlaPheSerIleValAlaLeuArgThrIleArgSerValSerArgIleValPheSerI 241
QY 696 TCTTCAACACCTGATCTGGTGTTCACCTGACGACGACAGACGAGACCTTATCTTAAGGC 755
Db 241 sLeuHisAspPheAspPheLeuArgTyrHisIleLeuSerArgGlyThrGlyIleuAsnArgI 261
QY 756 TATTGACAGAGAAACAGGGGATCAAGTTTGTGCTTAAGTGTGATTTAAATCTTTCT 815

```

```

Db 261 eIleAspArgGlySerArgAlaIleAsnPheIleLeuSerAlaMetValPheAsnValVa 281
QY 816 TCCATCATGTTTGAAGTGAATGCTTGTACGTGGTGTGTTGTTATATCAAAATGCGTGCCA 875
Db 281 lProThrIleLeuGlnIleSerMetValSerGlyIleLeuAlaTyrLysPheGlyAlaAl 301
QY 876 GTTTGCTTTGGTAAACCTTGGAAACACTTGTACATACAGACATTCACAGTTGCAGTGCAC 935
Db 301 aPheAlaTyrIleThrSerLeuSerValGlySerTyrIleValPheThrLeuAlaValTh 321
QY 936 ACGGTGAGAACTAGATTGAAATAGAACCAAGCAAGATTAATGATGACGATATGC 995
Db 321 rGlnTyrArgThrLysPheSerLysAlaMetAsnLysAlaAspAsnAspSerThrAr 341
QY 996 TGGTATAGACTCACTGCTGAATTAATGAACCTGGAAGTAAATTAATGAATGAATGA 1055
Db 341 gAlaIleIleAspSerLeuLeuIleAsnTyrGlnThrValLysTyrPheAsnAsnGlyArgI 361
QY 1056 AGCAGACAGATATGATGATTTTGAAGACGTATGACATGCTTCATTGAAAAGTACCTC 1115
Db 361 uAlaGlyLysTyrAspGlnPheLeuLysLysTyrGlyAspAlaAlaLeuGlnThrGlnAr 381
QY 1116 TACTCTGGCTATGCTGAACCTTGTCAAAAGTGCATATTTTACGTGCTGTTAACAGCTAT 1175
Db 381 gSerLeuAlaPheLeuAsnPheGlyGlnSerIleIlePheSerThrAlaLeuSerThrAl 401
QY 1176 AATGATGCTCGCAGTCAGGAAATGTGGCAGATCCCTTACCTTACCTTGTGAGATCATGAAT 1235
Db 401 aMetValLeuCySerGlnIleIleMetAsnGlyGlnMetThrValGlyLysPheValIle 421
QY 1236 GGTGAATGACCTGCTTTTTCAGCTTTTCATTACCCCTGAACTTTTGTGGAACTGATATAG 1295
Db 421 tValAsnGlyLeuLeuPheGlnIleuSerLeuProLeuAsnPheLeuGlySerValTyrAr 441
QY 1296 AGAGACTAGCAAGCACTCATATATGAACACCTTGTACTTACTTCAAGTAGACAC 1355
Db 441 gGlnThrIleGlnIleSerLeuValAspMetLysSerMetPheGlnLeuLeuGlnLysSe 461
QY 1356 CCAAAATTAAGCAAAAGTGAATGCTGCTCTCCCTTCAGATACACACACAGACGTAACGT 1415
Db 461 rAspIleThrAsnThrSerAspAlaLysProLeuValLeu-----LysGlyGlyAsnI 479
QY 1416 GGCCTTTGATTAATGTCATTTTGAATATCAATTAAGGGCCAGAAAGTCTTATAGTGAATATC 1475
Db 479 eGlnPheGlnAsnValHisPheSerTyrLeuProGlnArgLysIleLeuAspGlyIleSe 499
QY 1476 CTTTGAAGTCCCTGCGAGAAAGAAAGTGGCCATTTGAGAGGTAGTGGGTACAGGAAAG 1535
Db 499 rPheValIleProAlaGlyLysSerValAlaIleValGlyThrSerGlySerGlyLysSe 519
QY 1536 CACAATAGTAGGCTATTAATTTGCTTCTATAGGCTCAAAAGGTAAGCATTTATCTTGC 1595
Db 519 rThrIleLeuArgMetLeuPheThrPheAspThrAspSerGlyAsnIleArgIleAs 539
QY 1596 TGGTCAAAATATATCAAGATGTGAGCTGGAAGACCTTGGAGGGGAGTGGAGTGTATCC 1655
Db 539 pGlyGlnAspIleLysGlyValArgLeuAspSerLeuAspSerIleGlyValValPr 559
QY 1656 TCAAGATGCTGCTCTTCCATATATCTATTAACAACCTTTATATGAGAAACATCAG 1715
Db 559 oGlnAspThrValLeuPheAsnAspThrIlePheHisAsnIleHisTyrGlyArgLeuSe 579
QY 1716 TGGTCACTGAGGAGGTGATGACAGTGAAGAAATTAAGCTGAGCTTCATGATGCAATTC 1775
Db 579 rAlaThrGlyLeuGlnValTyrGlyAlaAlaArgArgAlaAlaIleHisGlnThrIleSe 599
QY 1776 TCGAATGCCACATGATATGACACCCAGTAGGGGACAGAGCATGAGCTTTCAGAGAG 1835
Db 599 rAsnPheProAspLysTyrSerThrIleValGlyIleArgGlyLeuLysLeuSerGlyI 619
QY 1836 AGAAAGCAAAAGATGCAATTCGAAAGACGCAATTTGAAAGACCCCACTACACTTA 1895
Db 619 yGlyLysGlnArgValAlaLeuAlaArgThrPheLeuLysSerProAlaIleLeuLeuCy 639

```

```

QY 1896 TGATGAAGCTACTTCATGCTAGATTCAGTAAGAGACTTATCTTGTCAGCATGAA 1955
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 639 SASPIULATHRSERIALALEUASPSTHRTHTGULAGULIELEUASALALEULY 659
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 1956 GGATGTGGTCAACACAGAACTTCTATTTCATTCGCACACAGATGTGCAACAGTGTGA 2015
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 659 SALALEUASERASNAARGTHRSERILEPHELLEALHISARGLEUTHRTHTRALAMEGL 679
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 2016 TGACATGAATTCATTCCTTGATCAGGCTAGGTAAGCCGACCGTACCCACCATGG 2075
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 679 NCYASPIULIEVALIEALGULASNGLYLSVALGULGULNGLYPROHISASPGI 699
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 2076 TTTCCTGCTAACCTCTCATAGTATCATTCAGAAATGTGCATACACAGAGACGCGTGT 2135
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 699 ULEULEUGLYLS---SERGLYARGTYRALAGLNEUTHRTHTGULNGINASNSERVA 718
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 2136 GCAGAACCATGAT 2148
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 718 LASPMELEUASP 722
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|

RESULT 4
QY 09M08 PRELIMINARY; PRT; 677 AA.
ID 09M08;
AC 09M08;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ABC transporter-like protein.
GN A1G28630.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Leonard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AL161573; CAB81451.1.-.
DR InterPro; IPR003593; AAA_Atpase.
DR InterPro; IPR001140; ABCtransportTM.
DR InterPro; IPR003439; ABC transporter.
DR Pfam; PF00664; ABC membrane; 1.
DR Pfam; PF00005; ABC trans; 1.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR ATP-binding; Transport.
KW SEQUENCE 677 AA; 75337 MW; 8D23CE38CDFEEA6 CRC64;

Alignment Scores:
Pred. No.: 1.02e-106 Length: 677
Score: 1657.50 Matches: 326
Percent Similarity: 74.30% Conservative: 128
Best Local Similarity: 53.36% Mismatches: 152
Query Match: 38.96% Indels: 5
DB: 10 Gaps: 4

AF133659 (1-2345) x 09M08 (1-677)
QY 316 ACAGACCCAAAGAGGGTTAAAGATGTTGATCTCGGAAATCATAAAGCAATGCTT 375
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 68 THIRPROAENGINSAPGINTHRYSRHLASERISLVSILIEUATHRTHTRILESER 87
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 376 TCTTATGTGTGGCCCAAGACAGCCAGATCTACAGCTAGAGTTCATTTGCTGGGA 435
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|

```

```

Db 88 SERTYRLEUTHRPMELYSASPASNPROGLIEUATHRPHEARVALLIEALALEUALA 107
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 436 TTTTGGGTGGTGAAGGCGCATGATATGTTGGTCTCTTCATGTTAAATATGCTGA 495
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 108 CYSLEULIEGLYALALYSRHELEUASNVALGLNALPROHEULEUPHLYSLEUSERILE 127
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 496 GACAGCCTCAACAGATGTGGGAAACATGTCGAACCTGAGTATGACCAATATACAGTT 555
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 128 ASPLEUUSERISERTYRSEUSERSERTHRILEHRASPERASN---PROTYRLEU 146
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 556 GCACACCATGGCAAA-----GCAGTTCTGATGGCTATGTTGATGATCAAGAGCTGAGCT 609
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 147 ALAIALPHEALATHRPROSERSERVALLEULIEGLYRISGLYILEALAAYSERGLYSER 166
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 610 GCTTTTAAACAGAGTTCGAAATGACGTATTTGGCAAGGTAGCCGAAATTCATCGGA 669
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 167 SERIALPHEASNGILUEALAGTHRALAVALPHESERISVALSERLEUALGTHRIEARG 186
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 670 AGAATGACCAAAATGCTCTTCCATCTTCACAACTGATCTGGGTTTCACCTGAGC 729
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 187 SERVALSERARGLYSVALLEUSERHISLEUHIASRPEUASPHEUARGTYRHISLEUASN 206
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 730 AGACAGACGGAGCTTATCTATAGAGCTATTGACAGAGAACAGGGTATCAGTTTGTG 789
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 207 ARGGLUTHRGLYALALEUASNARGILEILEASPARGLYSERAGALALIASRTHILE 226
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 790 CTGAGTGGTGTGATTTAAATCTCTCCATCATGTTTGAAGGATGCTGTGCAGTGT 849
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 227 LEUSERIALMETVALPHEASNAVALPRTHTRILEUENGLIISERMETVALTHNGLY 246
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 850 GTTTTGATTAACAATCGCGTCCAGTTCCTTGGTAAACCTTGACACATTTGTGACA 909
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 247 ILEUENALATYRASNPEGLYPROVALPHEALALEUILLHRSERLEUSERVALGLYSER 266
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 910 TACACAGCATTCACAGTTGCAGTCACACGSGTGAGACTGATTTAAATAGAAATGAC 969
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 267 TYRILEALPHEATHRLEUENVALTHNGLNYRTHRGLYSRHEARGVLSALAMEASN 286
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 970 AAACAGATATATGATGACGATATGCTGTATAGACTCACGTCGATATTAAGAACTGTG 1029
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 287 GLNALASPENASPALASERTHARGALALILEASPERLEUENALANTYRGLUTHVAL 306
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 1030 AAGTATTTTAAATATGAAGATATGAAGACAGAGATATGATGATTTTGAAGACGTAT 1089
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 307 LYSYRHEASNAENGLUASPTYRGLUALARGYSERYRASPHEULEUGLYARGTYR 326
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 1090 GAGACTGCTTATGAAGAAGTACTTACTCTGTGCTATGCTGAACCTTGGTCAAGTCT 1149
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 327 GLUASPALALALEUGLINTHRLGLNLYSERLEUHALPHELEUASPHEGLYGLNSERPHE 346
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 1150 ATTTCAGTGTGCGTTTACAGCATATAGTGCTGCGCCAGTCAAGGAATTTGCGCAGGT 1209
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 347 ILEPHESESRTHRALALEUSERTHSRMETVALLEUYSERGLNGLYILEMEASNGLY 366
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 1210 ACCCTTACTGTGGAGATCTAGTATGTAATGTAATGACTGCTTTTTCAGCTTCAATACC 1269
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 367 GLUMETHRVALGLYASPLEUVALMETVALASNGLYLEUENPHEGLINLEUSERLEUPRO 386
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 1270 CTGAACCTTTGCGGAAGCTGATATAGAGACTAGACAGACAGCTCATATAGAACACC 1329
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 387 LEUTYRPHLEUENGLYGLYVALTYRARGLUTHRVALGLNGLYLEUENALAPMETVLSER 406
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 1330 TTGTTTACTCTACCAAGTAGAGACACCAATTTAAAGACAAAGTAGGAGCATCCCTCT 1389
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 407 LEUPHEGLINLEUENGLUGLARGSERAPLIEGLYASPLYASAPTHRTHTGLUTHLYLEU 426
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 1390 ---CAGATCACACCAACAGACAGCTACCGTGGCTTTGATATATGCTATTTGAATACAT 1446
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 427 PROPROLEUVALLEUARGGLYSERILESRPHEGLIUSNALHISPHESERTYRLEU 446
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 1447 GAGGCGCAAGAAAGCTCTTATGGAATATCTTTGAAGTCCCTGACAGAAAGAAAGTGCC 1506
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 447 PROGLUTHRGLYSILEUENASPGLYLISERPHEGLIUALPROALAGLYLSERVALALA 466
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|

```





```

Qy 1210 ACCCTTACTGTTGAGATCTAGTAATGTAATGAGACTGCTTTTTCACGCTTTCATTACCC 1269
Db 368 GtmetThrValGlyAspLeuValMetValAsnGlyLeuLeuPheGlnLeuSerLeu 387
Qy 1270 CTGAACCTTCTGGGAACTGTATATAGAGACTAGCAACGACTCATATGATATGAAACC 1329
Db 368 LeuTyrPheLeuGlyGlyValTyrArgGlnThrValGlnGlyLeuValAspMetLysSer 407
Qy 1330 TTGTTTACTCTACTAGAGGTAGACCCCAATTAAGCAAGGTAGGACATCCCTCCCT 1389
Db 408 LeuPheGlnLeuLeuGlnGlyLysSerAspIleGlyAspLysAspThrGluThrLysLeu 427
Qy 1390 ---CAGATCACACACAGACAGACAGCTGCGCTTGTATATGTCATTTTGAATACAT 1446
Db 428 ProPheLeuValLeuArgGlyGlyLysSerIleSerPheGlnValHisPheSerTyrLeu 447
Qy 1447 GAGGCCCAAGAGTCTTATGTAATATCTTGAAGTCCCTGCGAGAAAGAAAGTGGCC 1506
Db 448 ProGluArgLysIleLeuAspGlyIleSerPheGlnValProAlaGlyLysSerValAla 467
Qy 1507 ATTGAGAGGATGAGGTGAGGTCAGGAAAGCAATAGTAGGCTATTTTGGCTTCTAT 1566
Db 468 IleValIglySerSerGlySerGlyLysSerThrIleLeuArgMetIlePheArgPhePhe 487
Qy 1567 GAGCTCAAAAGGTAGCATTTATCTTGTGCTGCAAAATATACAGATGAGCTGGAA 1626
Db 488 AspThrAspSerGlyAsnValArgIleAspGlyGlnAspIleGlyGlnValThrLeuGlu 507
Qy 1627 AGCCTTCGAGGGAGGAGTGGAGTGTACTCGATGAGATGCTGTCTTCCATATATCTAT 1686
Db 508 SerLeuArgSerCysIleGlyValValProGlnAspThrValLeuPheAsnAspThrIle 527
Qy 1687 TATTACAACCTTTTATGAGAAACATGCTGCTTCACTGAGAGAGATGATGCACTGGCA 1746
Db 528 PheHisAsnIleLeuTyrGlyAsnLeuSerAlaThrGlnGlnGlnValTyrAspAlaAla 547
Qy 1747 AAATTAGCTGATCATGATGATCAATTTCTCGATGCCCATGATGATGATGACACCAAGTA 1806
Db 548 ArgArgAlaValIleHisAspThrIleMetLysPheProAspLysTyrSerThrAlaVal 567
Qy 1807 GGGGAAACAGAGACTCAAGCTTTTCAGAGAGAGAAAGCAAGTATGCAATTGCAAGACC 1866
Db 568 GlyGluArgGlyLeuMetLeuSerGlyGlyGlyGlnArgValAlaLeuValaArgAla 587
Qy 1867 ATTTGAAGGAGACCCCACTTACTCTATGATGATGATGATGATGATGATGATGATGAT 1926
Db 588 PheLeuLysSerProAlaIleLeuLeuGlyAspGlyValAlaThrAsnAlaLeuAspSerLys 607
Qy 1927 ACTGAAGAGACTATCTTGGTGCATGAGATGATGATGATGATGATGATGATGATGATGAT 1986
Db 608 ThrGlnAlaGlnIleMetLysThrPheArgSerLeuAlaSerAsnArgThrCysIlePhe 627
Qy 1987 ATTGCACACAGATTGTCAACAGTGTGATGAGATGATGATGATGATGATGATGATGATGAT 2046
Db 628 IleAlaHisArgLeuThrThrAlaMetGlnCysAspGlnIleIleValMetClnLysGly 647
Qy 2047 AAGTAGCCGAAACGTGTACCAACCATGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 2106
Db 648 LysValValIleGlyLysGlyThrHisGlnValLeuLeu---GluLysSerCylArgTyrAla 666
Qy 2107 GAAATGTGCATACACAGAGACGCGTGTGAG 2139
Db 667 LysLeuTyrThrGlnGlnHisSerThrLeuGlu 677

```

```

GN AT428620.
OC Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Leonard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA Mewes H.W., Lemcke K., Mayer K.P.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Sanchez-Fernandez R., Mari S., Dancs A., Rea P.A.;
RL "Functional half-molecule ABC transporters from Arabidopsis: the ATM
RT family."
RL Submitted (JUL--2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL, AL161573; CAB1450.1; -.
DR EMBL, AF287698; AAC09828.1; -.
DR HSSP, P13569; INBD.
DR InterPro, IPR003593; AAA_ATPase.
DR InterPro, IPR001140; ABCtransporterTM.
DR InterPro, IPR003439; ABC_transportr.
DR Pfam, PF00664; ABC_membrane_1.
DR Pfam, PF00005; ABC_tran; 1.
DR ProDom, PD000006; ABC_transportr; 1.
DR SMART, SM00382; AAA; 1.
DR PROSITE, PS00211; ABC_TRANSPORTER; 1.
KM ATP-binding; Transport.
SQ SEQUENCE 680 AA; 76049 MW; 42169C4319FBE880 CRC64;

Alignment Scores:
Pred. No.: 3,73e-104 Length: 680
Score: 1620.50 Matches: 321
Percent Similarity: 71.32% Conservative: 129
Best Local Similarity: 50.87% Mismatches: 170
Query Match: 38.09% Indels: 11
DB: Gaps: 5

AF133659 (1-2345) x Q9M0G9 (1-680)
Qy 256 GTATGGCCATGATGAAAGAGACATGTTGCGATGCTCATGACAGAGACTTCAC 315
Db 57 IleThrProMetValaAsnAlaArgValMetPhePheSerThrSerThrSerAlaProHis 76
Qy 316 ACAAGCCCAAAAGAGGTTTAAAGATGTTGATCTCGGAAATCATMAAAGCAATGCTT 375
Db 77 -----ProGluLysIleLeuAsnArgThrSerSerGluAsnIleLeuArgMetIleSer 93
Qy 376 TCTTATGTTGTCGCCCAAGACAGGACAGATCTACGAGCTGAGTGGCATTTCCGCTGGGA 435
Db 94 SerTyrLeuTyrMetLysAspAsnProLysLeuLysPheArgValIleSerAlaPheAla 113
Qy 436 TTTTGGGTGTGCAAGGCCCATGAATATTGTGTTCCCTTCATGTTTAAATATGCTGTA 495
Db 114 CysLeuValGlyAlaLysPheLeuAsnValGlnValProPheLeuPheLysValAlaIle 133
Qy 496 GACAGCTCAACACAGATGTGGGAAACATGCTGAACCTGAGTATGCAACCAATACATT 555
Db 134 AspThrPheSerSerSerSer-----PheValAspSerAsnProTyrLeuVal 149
Qy 556 GCAACCATGGCAACA-----GCAGTTCTGATTTGGTATGCTGATGATGAGCTGGAGCT 609
Db 150 AlaAlaPheAlaThrProSerSerValLeuIleGlyTyrGlyIleAlaArgSerGlySer 169
Qy 610 GCTTTTTTTAAAGAGTTGGAATATGAGTATTTGGCAAGGTAGCCCAAAATTCAATCCGA 669
Db 170 SerAlaPheAsnGlnLeuArgThrSerValPheSerLysValAlaLeuArgThrIleArg 189

```

RESULT 6  
 Q9M0G9 PRELIMINARY; PRT; 680 AA.  
 AC Q9M0G9;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE ABC transporter-like protein (Half-molecule ABC transporter  
 DE ATW2).



QY 670 AGAATAGCCAAAATGCTTTTCTCCATCTTGCACAACTGGATCTGGTTTTACCTGAGC 729  
 Db 190 ThrilesSerArgLysValleuSerArgLeuHisAspLeuAspLeuArgTyHisleuAsn 209  
 QY 730 AGACAGACGGGAGCTTTATCTTAAGCTTTGACAGAGAACAGGGGTATGATTTTGTG 789  
 Db 210 ArgAspThrGlyAlaLeuAsnArgIleleAspArgGlySerArgAlaIleAsnThrIle 229  
 QY 790 CTGAGTGGCTTGTATTATCTTCTCCATCGATGTTGAAGTGAAGTGGCTGTCAGTGT 849  
 Db 230 LeuSerAlaMetValPheAsnIleMetProThrIleLeuGluIleSerMetValSerCys 249  
 QY 850 GTTTTGTATTAAATGCGGTGCCAGTTTGGCTTTGTAACCTTGGAAACCTTGATGTA 909  
 Db 250 IleleuAlaTyrlsPheGlyAlaValTyrlaIleuIleThrCysLeuSerValGlySer 269  
 QY 910 TACACAGATTCACAGTTGTCACACAGGTGAGAGAACTAGATTGAATGAATGAAC 969  
 Db 270 TyrlleAlaPheThrIleuAlaMetThrGlnTrpArgIleLysIleArgLysAlaMetAsn 289  
 QY 970 AAACAGATTAATGATGAGGTAAATGCTGTATGACTCAGCTGGAATTTAATGAACGTG 1029  
 Db 290 GluAlaGluAsnAspAlaSerThrArgAlaIleAspSerLeuIleAsnTyrlGluThrVal 309  
 QY 1030 AAGTATTTTATATGAAGAATATGAAGACACAGATATGATGATTTTGAACAGCTAT 1089  
 Db 310 LysTyrlPheAsnAsnGluAspTyrlGluAlaArgLysTyrlAspGlnLeuHisGluAlaAsnTyrl 329  
 QY 1090 GAGACTGCTTCAATGAAAGTACTCTACTGCTGCTATGCTGAACCTTGGTCAAGTCT 1149  
 Db 330 GluAspAlaIleuGlnSerArgLysSerPheAlaLeuAsnAsnPheGlnSerPhe 349  
 QY 1150 ATTTCAGTGGCTTAAACAGCATATATGCTGCTCGCAGTCCGGAATTTGTCGAGCT 1209  
 Db 350 IlePheSerThrAlaLeuSerThrAlaMetValLeuCysSerGlnGlyIleMetAsnGly 369  
 QY 1210 ACCCTACTGTTGAGATCTAGTATGATGATGATGATGATGATGATGATGATGATGAT 1269  
 Db 370 GlnMetThrValGlyAspLeuValMetValAsnGlyLeuLeuPheGlnLeuSerLeuPro 389  
 QY 1270 CTGAACCTTTCTGGGAACGTATATATAGAGACATGACAGACACTCATATATGACACC 1329  
 Db 390 LeuTyrlPheLeuGlyValIleValTyrlSerAspThrValGlnGlyLeuValAspMetLysSer 409  
 QY 1330 TTGTTTACTCTACTCAAGTGTGACACCAATTAAGACAAAGATGAGCATCTCCCTT 1389  
 Db 410 MetPheLysPheLeuGlnGluArgSerAspIleGlyAspLysAspIleAspArgLysLeu 429  
 QY 1390 ---CAGATCACACACACAGACAGTACCGTGGCTTTGATATATGATTTTGAATACATT 1446  
 Db 430 ProProleuValleuLysGlyGlySerIleSerPheGluAsnValHisPheSerTyrlLeu 449  
 QY 1447 GAGGGCCAGAAAGTCTTCTTGTAGTAATATCTTTGAAGTCCCTGAGAGAAAGAGGCC 1506  
 Db 450 ProGluArgLysIleLeuAspGlyIleSerPheGluValProAlaGlyLysSerValAla 469  
 QY 1507 ATTCTAGAGGTAGTGGTCTGAGGAGAAAGACATAGTGAAGCTATATTTGCTTCTAT 1566  
 Db 470 IleValIleSerSerGlySerGlyLysSerThrIleLeuArgMetIlePheArgPhePhe 489  
 QY 1567 GAGCCTCAAAAGAGTACATTTCTTGGCTGTCACAAATATACAAAGATGAGCCTGGAA 1626  
 Db 490 AspValAspSerGlyAsnValLysIleAspGlyIleAspGlyIleAsnValArgLeuGlu 509  
 QY 1627 AGCCTTGGAGGAGAGTGGAGTGTACCTCAGAGATCTGTCTTCTCAATATATATT 1686  
 Db 510 SerLeuArgSerSerIleGlyValIleValProGluAspThrValleuPheAsnAspThrIle 529  
 QY 1687 TATTACAACCTTTATATGAGAAATCAGTGTCTTCACTGAGAGAGTGTATGACGTGCA 1746  
 Db 530 PheHisAsnIleHisTyrlGlyAsnLeuSerAlaThrGlnGluGluValTyrlAsnAlaAla 549

QY 1747 AATTTAGCTGACTCATATGATGCAATTTCTTGAATGCCACATGGATATGACACCCAACTA 1806  
 Db 550 ArgArgAlaIleAlaIleHisAspThrIleMetLysPheProAspLysTyrlSerThrAlaVal 569  
 QY 1807 GGGGAACGAGAGACTCAAGCTTTTACGAGAGGAGAAAGAAAGATGATGATTTGCAAGGCC 1866  
 Db 570 GlyGluArgGlyLeuMetLeuSerGlyGlyGluLysGlnArgValAlaLeuAlaAspAla 589  
 QY 1867 ATTTTGAAGACCCCGCAGTATCTATATGATGAAGCTACTTATGATTTGATTTTCATT 1926  
 Db 590 PheLeuLysSerProAlaIleLeuLeuCysAspGluAlaThrSerAlaLeuAspSerLys 609  
 QY 1927 ACTGAAGACTATTTCTTGGTCCATGAGATGTGTGCTAAACAGACTTCTATTTC 1986  
 Db 610 ThrGluAlaGluIleMetLysThrLeuArgSerLeuAlaSerAsnArgThrCysIlePhe 629  
 QY 1987 ATTGCACACAGATTTGTCAACAGTGTGATGACAGATGAATCATTTGCTTGGATCAGGT 2046  
 Db 630 IleAlaHisArgLeuThrThrAlaMetGlnCysAspGluIleLeuValMetGluLysGly 649  
 QY 2047 AAGGTACCCGAGCGTGTATCCACCATGCTTGTGCTTAACCTCATATGATCTATTCA 2106  
 Db 650 LysValValGluLysGlyThrHisGluValLeuLeuGlyLys---SerGlyArgTyrlAla 668  
 QY 2107 GAATGTGCGATACACAGACAGCCGCTGTGACG 2139  
 Db 669 LysLeuTrpThrGlnGlnAsnSerLysLeuGlu 679

## RESULT 7

Q8T9W2 PRELIMINARY; PRT; 697 AA.  
 ID Q8T9W2;  
 AC Q8T9W2;  
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
 DE ABC transporter ABCB5.  
 GN ABCB5.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Eukaryota; Dictyostelid; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RA Anjard C., Loomis W.F.;  
 RT "Evolution of the ABC transporters of Dictyostelium."  
 RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF466308; AAL74252.1; -  
 SQ SEQUENCE 697 AA; 78531 MW; 304C7BC2F344E2F CRC64;

## Alignment Scores:

| Pred. No.:             | 1,166-102 | Length:       | 697 |
|------------------------|-----------|---------------|-----|
| Score:                 | 1599.00   | Matches:      | 312 |
| Percent Similarity:    | 72.02%    | Conservative: | 123 |
| Best Local Similarity: | 51.66%    | Mismatches:   | 153 |
| Query Match:           | 37.59%    | Indels:       | 16  |
| DB:                    | 5         | Gaps:         | 5   |

AF133659 (1-2345) x Q8T9W2 (1-697)

QY 316 ACAGACCCAAAGAGAGGTTAAAGATGTTGATCTCGGAAATCATTAAGCAATGCTT 375  
 Db 108 ThrAspIleLysGlu-----ThrGlnAsnSerThrMetSerLeuLysThrValPhe 125  
 QY 376 TCTTATGTGTGGCCCAAGACAGCCAGATCTACGAGTCTGAGTGTGCGATTTGGCTGGCA 435  
 Db 126 LysTyrlLeuTrpProLysAspAsnAspSerLysIleArgIleIleThrSerValLeu 145  
 QY 436 TTTTGGGTGTGCAAGGCCAATGATTTGTTGCTTCCATGTTTAATATGCTGTA 495  
 Db 146 LeuLeuLeuSerAlaLysValleuThrValGlnIleProHeIlePheLysAspIleVal 165  
 QY 496 GACAGCTCAACAGATGTGCGGAAACATGCTGAACCTGAGTATGACCAATAATACATT 555

|    |      |  |      |
|----|------|--|------|
| Db | 166  | AspserleuthrThrInglInserGluMetLeu-----                             | 176  |
| Oy | 556  | GCACACATGSCACACAGCATTTGATTTGGCTATGGTATCAAGACTGACCTGCTTT            | 615  |
| Db | 177  | ---ThrlleuProLeuGlyLeuLeuLeuAlaTyGlyAlaValAlaIleThrSerAsnGly       | 195  |
| Oy | 616  | TTTAAACGAAGTTGCAAAATGCAAGTATTTGGACAGTGGCCCAAAATTCAATCCGAAGATA      | 675  |
| Db | 196  | PheGlnGluLeuAspArgGlnThrThrIlePheSerLysValAlaIleAspAlaIleArgAspVal | 215  |
| Oy | 676  | GCCAAAATATGCTCTTCTCCATCTTCCACACCTGGATCTGGTGGTTTTCACCTGACGACAG      | 735  |
| Db | 216  | SerCysSerThrPheLysValArgLeuHisGlnLeuAspLeuThrPheHisLeuSerArgGln    | 235  |
| Oy | 736  | ACGGAGACTTTATCTTAAGCTATTGACACAGAAACAAGGGATATCACTTTGTCTCTGAGT       | 795  |
| Db | 236  | ThrGlySerLeuSerAspArgIleIleAspArgGlyGlyArgGlyLysIleAsnPheLeuAsn    | 255  |
| Oy | 796  | GCTTTGGATTTAATCTTCTTCCATCATGATTTTGAAGTCAATGCTTGTACAGTGGTCTTTTG     | 855  |
| Db | 256  | SerIleLeuPheHisValAlaProThrAlaPheGlnIleSerLeuValSerLysValMet       | 275  |
| Oy | 856  | TATTACAATACGGTGGCCAGATTTGGCTTTGGTGGACCTTGGAAACCTTGGTCAACACACA      | 915  |
| Db | 276  | TyrThrThrLeuGlyLysPheGlyLysSerAlaLeuSerLeuAlaThrIleAlaIleArgThr    | 295  |
| Oy | 916  | GCATTCACAGTTGCAGTCAACGGTGGAGAACTAGATTTTGAATAGAAATGAACAACACA        | 975  |
| Db | 296  | ValPheThrValLysValThrLysLysTyrArgThrGlnPheArgValLysMetAsnLysMet    | 315  |
| Oy | 976  | GATATATGATGCAAGTAATGCTGCTATAGTCACTGCTGAATTAATGAACCTGTAAGAT         | 1035 |
| Db | 316  | AspAsnGlnLysAspLysMetLeuAspSerLeuIleAsnGlnThrValLysTyr             | 335  |
| Oy | 1036 | TTTAAATATGAAAGTATGAAAGCACAAGATATGATGATTTTGAAGACGTATGAGACT          | 1095 |
| Db | 336  | PheAsnAsnAspAlaLeuGluValGluArgTyrHisAsnTyrLeuLysGluTyrAspLys       | 355  |
| Oy | 1096 | GCTTCATTGAAAGTACCTCTTACTCTGGCTATGCTGAACCTTGGTCAAGTCTATTTTC         | 1155 |
| Db | 356  | AlaSerLeuLysThrThrSerSerLeuSerPheLeuAsnPheGlyGlnAlaLeuIlePhe       | 375  |
| Oy | 1156 | AGTTCGCTTTAAACGCTATTAATGGTCTGCCAGTCAAGGAAATTGTGGCAGTACCTT          | 1215 |
| Db | 376  | SerLeuSerMetThrAlaMetMetIleMetAlaGlnGlyValAlaGlnLysAsnLeu          | 395  |
| Oy | 1216 | ACTGTGGAGATCTAGTAAATGCTGAATGAACGTGTTTTTACGTTTCATTACCCCTGAC         | 1275 |
| Db | 396  | SerValGlyAspLeuValLeuValAsnGlyLeuLeuPheGlnIleSerLeuProLeuAsn       | 415  |
| Oy | 1276 | TTTCTGGGAAGCTGATATATAGAGACTAGACAAGCACTCATATAGATGAACCTTGT           | 1335 |
| Db | 416  | PheLeuGlyThrValTyrArgGlnIleLysGlnSerLeuValAspMetAspHisLeuPhe       | 435  |
| Oy | 1336 | ACTGTACTCAAGTATAGACACCCAAATTAAAGACAAAGTATGGCATCTCCCTTCAGATC        | 1395 |
| Db | 436  | SerLeuLeuAsnLeuAsnProLysIleSerAspAsnLysAspSerLysProLeuLysLeu       | 455  |
| Oy | 1396 | ACACACAGACAGCGATCCGCTGGCTTTGATTAATGATGATTTTAAATACATTAAGGGCCAG      | 1455 |
| Db | 456  | -----GluAsnGlyThrIleValAlaPheArgAspIleSerPheLysTyrIleAsnAspSerVal  | 473  |
| Oy | 1456 | AAAGCTCTTATGCAATATCTTTGAAGTCCCTGCACAGAAAGAAAGTGGCCCATTTGAGA        | 1515 |
| Db | 474  | GlnValLeuAsnAsnValSerPheGlnLysGlnGlyLysArgGlyIleAlaIleValGly       | 493  |
| Oy | 1516 | GGTATGGGTGACGAGAAAAGACAAATAGTAGGCTATTTATTTGCTTTCTATAGACCTCA        | 1575 |
| Db | 494  | SerSerGlySerGlyLysSerThrLeuLeuArgLeuLeuTyrArgPheTyrAspValSer       | 513  |
| Oy | 1576 | AAAGGTACCATTTATCTTTCGTCGTCGCAAAATATACAGATGTAGCCTGGAAAGCCTTGG       | 1633 |
| Db | 514  | SerGlySerIleGlnIleAspGlyGlnAspGlyLeuArgGlyLysIleGlnLeuGlnSerLeuArg | 533  |

|                   |  |   |              |
|-------------------|--|---|--------------|
| Qy                | 1636   | AGGCAGCTGGAGAGGTGACCTCAGAGATGCTGTCCTTCCTCAATAATCAATATTATTAAC      | 1695         |
| Db                | 534  | Lyshist1leg1valValProGlnaapThrValLeupheAenAepThr11e1ytr1rsh       | 553          |
| Qy                | 1666   | CTCTTAATGGAAMAACATCACTGCTTCACTGAGGAAGTATATGACGTGGCAAAATTAGCT      | 1755         |
| Db                | 554  | 11e1a1ytrGlyshenProaenA1a1ThrLysG1uG1n1valG1uAaenA1a1a1a1a        | 573          |
| Qy                | 1756   | GGACTTCATGATGCATTTCTTCGATATGCCACATGCATATGACATCCCAAGTGGGGACGA      | 1815         |
| Db                | 574  | His11e1h1g1n1val11e1LeuAenMeLysAaenGlyTyraSpThrValValG1yG1uAr     | 593          |
| Qy                | 1816   | GGACTCAAGCTTCAGAGAGAGAAAAGCAAGAGTAGTATGCAAGCCATTTTGAAG            | 1875         |
| Db                | 594  | G1yLeuLysLeuSerG1yG1yG1uLysG1nA1yValSer11e1a1ArA1a11e1uLys        | 613          |
| Qy                | 1876   | GACCCCCAGCTCACTCATATGATGAAGCTACTTCATCGTTAGATTGACTTAAGAG           | 1935         |
| Db                | 614  | AaepSerPro11e1valPheTyraSpG1uA1a1ThrSerSerLeuAaSpThrG1uLysG1uLys  | 633          |
| Qy                | 1936   | ACTATTTCTGGTGCATGAAGATGTGGTCAACAACACAGAACTTCATTTTATTCGCAC         | 1995         |
| Db                | 634  | Leu11e1MetAaSpA1a1LeuArG1uLeuPheTyraG1yArG1yThrTh11e1Met11e1A1h1s | 653          |
| Qy                | 1996   | AGATTGCAACAGGTGTTGATGCAGATGAATCATGTCTTGATCAGAGGTAAAG---GTA        | 2052         |
| Db                | 654  | ArgLeuSerThr11e1ValAaPa1a1a1a1a1a11e11e1ValLeuG1yThrG1yG1y11e1le  | 673          |
| Qy                | 2053   | GCCGAACGTGTGATCCCAACCATGTTGTTGCTGCTAAACCTCATATGATATTCAGAAATG      | 2112         |
| Db                | 674  | LeuG1uArG1yAaenH1sG1nG1n1LeuLeu---G1uLeuG1uG1yLysTyraArgSerMet    | 692          |
| Qy                | 2113   | TGGCATACACG 2124  |              |
| Db                | 693  | TrpLeuA1aG1n 696  |              |
| RESULT 8          |  |   |              |
| Q9SNUS            |  |   |              |
| ID                | Q9SNUS   | PRELIMINARY;  | PRT; 936 AA. |
| AC                | Q9SNUS;  |   |              |
| DT                | 01-MAY-2000 (TrEMBLrel. 13, Created)                               |   |              |
| DT                | 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)                  |   |              |
| DT                | 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)                |   |              |
| DE                | ESTs A0067992 (C11433).  |   |              |
| OS                | Oryza sativa (Rice)  |   |              |
| OC                | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; |   |              |
| OC                | Spermatopsida; Magnoliophyta; Liliopsida; Poales; Poaceae;         |   |              |
| OC                | Ehrhartoideae; Oryzaceae; Oryza.                                   |   |              |
| OX                | NCBI_TaxID=4530;   |   |              |
| RN                | [1]  |   |              |
| RP                | SEQUENCE FROM N.A.   |   |              |
| RC                | STRAIN=CV. NIPONBARE;  |   |              |
| RA                | Sasaki T., Matsumoto T., Yamamoto K.;                              |   |              |
| RT                | "Oryza sativa niponbare (GAJ) genomic DNA, chromosome 6, PAC       |   |              |
| RT                | clone: P0538C01."  |   |              |
| RL                | Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.            |   |              |
| CC                | -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.             |   |              |
| DR                | EMBL; AP000391; BA88352.1; -                                       |   |              |
| DR                | InterPro; IPR003593; AAA_A1rase.                                   |   |              |
| DR                | InterPro; IPR001140; ABCTransportTM.                               |   |              |
| DR                | InterPro; IPR004339; ABC_transportr.                               |   |              |
| DR                | Pfam; PF00664; ABC_membrane; 1.                                    |   |              |
| DR                | Pfam; PF00005; ABC_tran; 1.  |   |              |
| DR                | ProDom; PD000006; ABC_transportr; 1.                               |   |              |
| DR                | SMART; SM00382; AAA_1.   |   |              |
| DR                | PROSITE; PS00211; ABC_TRANSPORTER; 1.                              |   |              |
| KW                | ATP-binding; transport.  |   |              |
| SEQ               | SEQUENCE 936 AA; 101685 MW; C52EE4AA5303AEEC CRC64;                |   |              |
| Alignment Scores: |  |   |              |
| Pred. No.:        | 3,73e-101  | Length:   | 936          |
| Score:            | 1577.50  | Matches:  | 360          |

Percent Similarity: 51.83%  
Beet Local Similarity: 38.79%  
Query Match: 37.08%  
DB: 10  
Conservative: 121  
Mismatches: 186  
Indels: 261  
Gaps: 12

AF133659 (1-2345) x Q95N05 (1-936)

```
QY 34 GCGGCGCGGCGCTGTTTGGAAAAGCCGCGACCTCCGCGATTCTGATCCGCTTTA 93
DB 27 AAAAAAAAAAProIlePhe-----ArgArgProProthrrValProArgProLeu 43
QY 94 GTCTCT-----GTTACGGCTCAGTCCGCGAGTGGAGCCACATCAACTCGGCGCTTG 147
DB 44 ProSerProLeuLeuGlyGlyPheGlyProAsn----- 54
QY 148 GGAACCGCTCGAGCCTACAGATTCCAGAGTCATTAAAGATACATGCGCAGAGATTG 207
DB 55 -----CysThrValTyrPro 59
QY 208 GGAAGAGCAATTCAGACAGATTG-----TTAGATCTCGAAGGCTCTCCAGTATGG 261
DB 60 GlyAspGlyLysTyrAlaProPheGlyArgLeuSerCysPheMetSerAspSerThrTyr 79
QY 262 CCACGTGATGAAAAAGAGACAGATGTTGGCATGCTCATGCA----- 300
DB 80 ProProProArgAspVal-----ArgGlyHisAlaPheSerThrSerAlaAsnAla 97
QY 301 ---GAGAGAGACTCCACAGACCCAAAGAA-----GGGTTAAAGATGTT 345
DB 98 ValAlaValGlyLysSerSerAspAspLysValLysLysAspLysSerLysLysAspVal 117
QY 346 GATPACTCG-----AAATCATAAAGCAATGCTTCTTATGTTGTCGCCCAA 393
DB 118 AspAspGlnIleAlaAspThrGlnIleLeuLysAsnLeuGlyLysTyrLeuLeuAsn 137
QY 394 GACAGGCGCAATCTACAGAGCTAGAGTTGCCATTTCGCTGGGATTTGGGTGGCAAG 453
DB 138 AspSerProAspPheArgPheArgLeuIleLeuSerLeuGlyLeuValGlyAlaLys 157
QY 454 GCCATGATATTTGTTGCTCCCTTCAATGTTAAATATGCTGAGACAGCTCCACAGATTG 513
DB 158 ValIleAsnValGlnValProPheLeuPheLysLeuValAlaAspTrpLeuAlaLeu 177
QY 514 TCGGGAACATGCTGACCTGAGTGATGACCAAAATACAGTTGCAACCATG----- 564
DB 178 AlaglyAlaGluThrSerLeuAlaSerPheThrGluAlaAsnAlaThrLeuAlaLeu 197
QY 565 -----GCACACAGCATTTCTGATTGGCTGATGCTGATCAAGAGCTGAGCTCTTTT 615
DB 198 PheAlaSerProAlaAlaValLeuIleGlyTyrGlyIleAlaArgSerGlyValSerAla 217
QY 616 TTTAACGAGTTGCGAATGAGATGATTTGGCAAGGTACCCAGAAATTCAAATCCAGAAATA 675
DB 218 CysThrIleuLeuArgAsnAlaValPheSerLysValThrLeuArgAlaIleArgSerVal 237
QY 676 GCCAAA----- 681
DB 238 SerMetProSerArgGlyArgGlyThrArgArgArgProAspLeuArgGlyProAlaAla 257
QY 681 ----- 681
DB 258 AlaLeuAlaLysValAlaAlaAlaValAlaAspGluAlaGluValAlaGluPheValGlu 277
QY 681 ----- 681
DB 278 MetAlaValThrValLeuValGluArgArgTrpArgSerArgMetSerGlyGlyThrArg 297
QY 681 ----- 681
DB 298 GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyAlaAlaLeuValAspAlaArgArgTrpTrp 317
QY 681 ----- 681
```

```
DB 318 SerArgAlaSerLysGlyAlaHisGlyGlyGlyGlyGlyArgGlyArgGlyGly 337
QY 681 ----- 681
DB 338 GlyAlaGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyValGluValAlaAsp 357
QY 681 ----- 681
DB 358 GluAlaAlaThrLeuValGlyAlaArgArgSerSerArgArgTrpAspGluPheValArg 377
QY 681 ----- 681
DB 378 GlnPheLeuAlaAspArgLeuValProAspLeuArgGlyIlePheLeuSerArgAspGln 397
QY 681 ----- 681
DB 398 ProIleProProthrrProLysProAsnThrProLysSerGlyPheValProSerHisPro 417
QY 681 ----- 681
DB 418 IleProSerLeuGlnProAsnThrThrLeuLeuValLysLeuThrValIleLeuProSer 437
QY 682 -----AATGCTTTCTCCATCTTCACAACTGGATCTGGGTTTCACCTG 726
DB 438 GluLeuLeuLeuLeuGlnValPheSerHisLeuHisGluLeuAspLeuArgTyrHisLeu 457
QY 727 AGCAGACAGACGAGCTTATCTAAGGCTATGAGCTATGACAGAGAACAGGCGTATCAGTTT 786
DB 458 SerArgGlnThrIleAlaLeuAsnArgIleIleAspArgGlySerArgAlaIleAsnTyr 477
QY 787 GTCCTGAGCTTTGGTATTAATCTTCTCCATCATGTTGAAGTATGATGATCTTGCAGT 846
DB 478 IleLeuThrValMetValPheAsnValAlaProThrIleLeuGlnIleGlyMetValSer 497
QY 847 GCGTTTGTGATTAACAATCCGCTGCCAGTTGCTTGGTAACTCTGGAACACTTGC 906
DB 498 SerIleLeuAlaTyrLysPheGlySerThrPheAlaTrpIleHisSerValAla 517
QY 907 ACATACACAGCATTCACAGTGCATCACACGCGTGAGAGAACTGATTTAGAAATAGAATG 966
DB 518 ThrTyrIleAlaPheThrLeuAlaValThrGlnThrArgThrLysPheArgThrAlaMet 537
QY 967 AACAAAGCAGATTAATGATCAGGATTAATGCTGATAGACTCACTGCTGAATTAAGAACT 1026
DB 538 AsnLysAlaAspAsnAlaSerSerThrValAlaValAspSerLeuLeuSerLysLys 557
QY 1027 GTGAGATTTTATATGAAGATATGAAGCACACAGATATGATGATTTTGAAGACG 1086
DB 558 ValLysTyrPheAsnAsnGlnGlnPheGluValGluLysTyrAspLysTyrLeuLys 577
QY 1087 TATGAGACTGCTTCAATGAAGTACTGATCTGCTGCTATGCTGAACCTTGTCAAGT 1146
DB 578 TyrGluAspAlaAlaLeuLysThrGlnSerSerLeuAlaTyrLeuAsnPheGlyGlnAsn 597
QY 1147 GCTATTTTCAAGTCCGTTTAAACAGCTATATAGTCTGCTGCCAGTACAGGAAATGTGGCA 1206
DB 598 IleIlePheSerSerAlaLeuSerThrAlaMetValLeuSerSerLysGlyValMetSer 617
QY 1207 GGTACCTTACTGTTGAGATCTAGTAATGTAATGAGACTGCTTTTTCAGCTTTCAATTA 1266
DB 618 GlyAlaLeuThrValGlyAspLeuValMetValAsnGlyLeuLeuPheGlnLeuSerLeu 637
QY 1267 CCCCTGAACTTCTGGGAAGCTGATATATAGAGACATGACAGACACTCATAGATATGAAC 1326
DB 638 ProLeuAsnPheLeuGlySerValTyrArgGluSerArgLysSerLeuIleAspMetLys 657
QY 1327 ACCTGTTTACTCTACTACAGATGAGACACCAATTTAAAGACAAAGATGATGATCTCC 1386
DB 658 SerMetPheGlnLeuLeuGlnGlyLysProGlyIleLysAspGluProHisAlaGlnPro 677
QY 1387 CTTGAGATCACACACAGACAGCTACCGGCTTTGATTAATGCAATTTTGAATACAT 1446
DB 678 LeuGlnPhe-----LysGlyGlyArgIleGluPheGluAsnValHisPheGlyTyrVal 695
```

```

Qy 1447 GAGGCCAGAAAGCTTATAGTGAATATCTTGAAGTCCCTGACGAAAGAAAGTGGCC 1506
Db 696 ProgluAglvIsIleuLeuYsGlYalArhRheMhValProIaGlYlYsSerVala1a 715
Qy 1507 ATTGTAGAGAGTAGTGGCTCAGGAGAAAGCACAATAGTAGAGGCTTATTTGCTTCTAT 1566
Db 716 IleValGIYthnSerGIYSerGIYsSerThrIleLeuArGIleuLeuPheArGhePhe 735
Qy 1567 GAGCCTCAAAAGGAGTAGATTTATCTGCTGCAAAATATPACAAGATGAGCCTGAA 1626
Db 736 AsperSerSergIYSerIleArGIleArGIYlAspGIYlAspIleArGIYlValThrLeuAsp 755
Qy 1627 AGCCTTCGAGGAGGAGTGGAGTGGTACCTCAGATGCGTCTCTCCATTAATCTATT 1686
Db 756 SerIeulArGIYsCyelIleGIYalValProGIlnAspThrValLeuPheAsnAspThrIle 775
Qy 1687 TATTAACAACCTTATATATGAAACATCAGTGTCTCACCCTGAGAAAGTATGACGTGCA 1746
Db 776 LysHIsaenIleGIlnTYrGIYArGIleuSerAlaThrArGIYlValTYrAspVala1a 795
Qy 1747 AAATTAAGTGCATTCATGATGCATTTCTTGAATGCCACATGATATGACCCCAAGTA 1806
Db 796 ArGaTgAlaAlaIleHIsaSprHrIleMetAspPheProAspLYrAsnThrVala1a 815
Qy 1807 GAGGAAAGAGAGACTCAGCTTTGAGAGAGAGAAAGCAAGAGATGACATTCAGAGAGCC 1866
Db 816 GIYlulArGIYlYleuLysSerGIYlYlYsGIlnArGIYalSerIleAlaArGIYal 835
Qy 1867 ATTTGAAGAGACCCCGCATCTACTATGATGAAAGTACTTCAATCGTTAGATTGATT 1926
Db 836 PheLysLysGIYlProSerIleLeuLeuCYsAspGIYlAlaThrSerAlaLeuAspSerThr 855
Qy 1927 ACTGAAGAGACTATCTTGTGTCCATGAAAGATGTGTGTAACAGAACTCTATTTC 1986
Db 856 ThrGIYlAserIleLeuAsnSerLeuLysThrLeuSerValAspArGIYHsrIlePhe 875
Qy 1987 ATTGCACACAGATTGTCAACAGTGGTATGATGCAGATGAATCATGTTGTTGAGGAT 2046
Db 876 IleAlaHIsaRGIleuThrThrAlaMetGIYsAspGIYlIleIleValLeuGIlnsngly 895
Qy 2047 AAGTAGCCGAAAGTGTGACCCACATGGTGTGTTGCTTAACCTCATATGATCTATTCA 2106
Db 896 GIuValIleGIYlGIYnGIYlProHIsaPheLeuLeuSerLYs---GIYGIYArGIYAla 914
Qy 2107 GAAATGTGCATACACAGAGAGC 2130
Db 915 GIuLeuTYrSerGIYnGIYAsn 922

RESULT 9
Q82D10 PRELIMINARY; PRT; 631 AA.
ID Q82D10
AC Q82D10
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Putative ABC transport protein.
GN YP02588.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Pankiewicz M., Sebailia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Baaham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;

```

```

RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ14152; CAC91388.1; *.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001140; ABC_transport.
DR InterPro; IPR003439; ABC_transport.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transport; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 631 AA; 71336 MW; 231DA74D67B64F3B CRC64;

Alignment Scores:
Pred. No.: 1,18e-92 Length: 631
Score: 1454.50 Matches: 293
Percent Similarity: 67.73% Conservative: 110
Best Local Similarity: 49.24% Mismatches: 179
Query Match: 34.19% Indels: 13
DB: Gaps: 3

AF133659 (1-2345) x Q82D10 (1-631)
Qy 340 GATGTGATCTCGGAAATCATAAAGCAATGCTTTTATGTGTGCGCCAAAGACAG 399
Db 35 AsnMetAspArGIYSerArGIleuLysPheLeuLeuProTYrLeuTYrProLYsAspAsn 54
Qy 400 CCAGATTCACAGAGCTAGAGTTCATTTGCTGGAGATTTTGGGTGGTGCAAAGCCATG 459
Db 55 ProLYsLeuArGIYTYrTYrLeuIleIleAlaLeuPhePheMetValIleAlaLYsSer 74
Qy 460 AATATGTGTTCCCTCATGTTTAAATATGCTGTAGACAGCTCAACAGATTCGCGGA 519
Db 75 ThrThrLeuValProLeuAlaTYrLYsSerMetValAspThrLeuSerSerGIYAsnAla 94
Qy 520 AACATGCTGAACCTGAGTATGACCAAAATACAGTTGCAACATGCAACAGCATTTCTG 579
Db 95 LysMetLeu-----AlaIleProIleSerLeuIle 104
Qy 580 ATTGGCTATGTGTATCAAGAGCTGAGCTGCTTTTAAACGAAGTTTCAATGCAAGTA 639
Db 105 IleAlaTYrGIYValAlaArGIYalSerAlaSerLeuLeuPheGIYlGIYLeuArGIYalMet 124
Qy 640 TTGGCAAGTAGCCGAAATCAATCGAAGATGACCAAAATGCTTCTCTCATCTT 699
Db 125 PheValHIsaLYsSerGIYAsnAlaThrArGIleuLeuGIYleuArGIYalPheArGIYlLeu 144
Qy 700 CACAACCTGATCTGGTGTTCACCTGAGACAGACAGACGAGCTTATCTTAAGCTATT 759
Db 145 HIsaLeuLeuSerLeuArGIYpHeHIsaLeuGIYnArGIYlThrGIYlYleuSerLeuSerIle 164
Qy 760 GACAGAGGAACAAGGGGATACAGTTTGTCTGAGTGTGTTGATTTTAATCTTCTTCCC 819
Db 165 GIuArGIYlTYrGIYnAlaValaLYsSerThYrValLeuSerArGIYleuLeuPheSerIleuPro 184
Qy 820 ATCATGTTGAAGATGAGCTGTGCTGAGTGTGTGTGATTAACAATGGCGGCGCCAGATT 879
Db 185 IleuPheGIYlIleThrLeuValSerValIleMetIleThrArGIYleuLeuSerGIYTYrPhe 204
Qy 880 GCTTGTGTAACCTTGGAAACCTGTGATACATACACAGATTCACAGTTGACAGTACACAGG 939
Db 205 AlaLeuAlaIleLeuValThrValSerCYsTYrIleLeuPheThrValMetAlaValGIY 224
Qy 940 TGGAGACTAGATTTAGATATGAATGAACAAGCAGATTAATGTCAGAGGTAATGCTGCT 999
Db 225 TYrArGIYlArGIYpHeArGIYlGIYnAsnLYsAlaAsnAlaAspAlaAsnThrLYsSer 244
Qy 1000 ATAGACTACAGCTGAATTAAGAAGCTGAAGATTTTAATTAATGAAGATATGAAGCA 1059
Db 245 IleAspSerLeuLeuAsnTYrGIYnThYrValLYsTYrPheGIYnAsnGIYAspPheGIYAla 264
Qy 1060 CAGAGATATGATGATTTTGAAGACGTATGAGACTGCTTCATTGAAGAAAGTACTTACT 1119

```

|   |   |      |                        |                               |                           |                 |        |      |
|---|---|------|------------------------|-------------------------------|---------------------------|-----------------|--------|------|
| D | b | 265  | GUUAGPhenAlenUserA     | gGlnLeuTyGlnU                 | YAlaAlaLeu                | LeuSerIn        | LeuSer | 284  |
|   |   |      |                        |                               |                           |                 |        |      |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1120 | CTGGCACTGAACTTTGGTCAAA | TGCTATTTCAGTGTGCGTTAA         | CAGCTAAATG                |                 |        | 1179 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 285  | PhenThrValLeuSerPhe    | GlnThrAlaIle                  | IleSerValGly              | LeuIleVal       | LeuMet | 304  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1180 | GTGCTGCCCATCAGGGAAT    | TGTGGAGGTAACCTTACTCTTG        | TGGAGATTAAGTG             |                 |        | 1239 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 305  | AlaMetAlaIaGlnGlyIle   | ValaGlnGlyGlnMet              | ThrIleGlyAsp              | PheValLeuVal    |        | 324  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1240 | AATGCACTGCTTTTACGCTT   | TCAATTCACCTTGAACTTTCTGGAA     | CTGTAAATAGAG              |                 |        | 1299 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 325  | AsnAlaTyrlLeuLeuGln    | LeuTyrlGlnProLeu              | AsnPheGlyPheIle           | TyrlSerGln      |        | 344  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1300 | ACTAGACAAGCACTCATAG    | TATGAACACCTTGTTACTTACTTA      | CTCAAGGTAGACACCA          |                 |        | 1359 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 345  | IleArgGlnAlaLeuIle     | AsnMetGlnAsnMet               | LeuAspLeuMet              | ValLysGlnGln    |        | 364  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1360 | ATTAAAGCAAAAGTAGGCA    | TCTCCCTTCAGATCAACACACAGAC     | AGTACCTGTC                |                 |        | 1419 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 365  | IleThrAspArgProAsp     | AlaLeuProLeuGlnLeu            | Thr-----                  | LysGlyGlnValArg |        | 382  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1420 | TTTGATATGTCATTTGAAT    | CATTGATGAGGCGAAGAAAGCTCTAG    | TGGAAATATTCCTT            |                 |        | 1479 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 383  | PheAspAlaValSerPhe     | SerTyrlAspProArgArgProIle     | LeuAsnAsnValSerPhe        |                 |        | 402  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1480 | GAAGCCCTGCAGAGAAAG     | AAAGTGCCCATTTGTAGAGAGTATGGG   | TACAGGAGAAAGACA           |                 |        | 1539 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 403  | ThrIleProProGlyLys     | ThrValAlaIleValGlyAla         | SerGlyAlaGlyLysSer        | Thr             |        | 422  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1540 | ATAGTAGGCTATTATTGCT    | TCTGTATGAGCCCTCAAAAGGGTAGCA   | TTTATCTTGCTGT             |                 |        | 1599 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 423  | LeuAlaArgLeuLeuPhe     | AsnArgPheTyrlAspValThrAlaGly  | ValAlaTyrlLeuAsp          |                 |        | 442  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1600 | CAAAATATACAAAGTGTAG    | CCCTGGAAAGCTTCGAGGAGCAGTGGAG  | TGTACTTACAG               |                 |        | 1639 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 443  | GlnAspIleArgGlyVal     | IThrGlnSerSerLeu              | ArgGlnAlaIleGlyIleVal     | ProGln          |        | 462  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1660 | GATGCTGCTCCCTCCAA      | TAATACTATTATTAACCTCTGTATG     | GAAGAAACATCACTGTCT        |                 |        | 1719 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 463  | AspThrValLeuPhe        | AsnAspThrLeu                  | ArgValAsnIleGlyTyrlGlyLys | ThrAspSer       |        | 482  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1720 | TCACCTGAGAAAGTAT       | TGAGAGTGGCAAAATTTAGCTGAGCACTT | CATGATGATCTTGTGA          |                 |        | 1779 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 483  | ThrAspGlnGlnIleGln     | ArgAlaIaIaLysLeuAlaIaIaIle    | IleIleGlnPheIle           | LeuSer          |        | 502  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1780 | ATGGCACTGATATGAC       | ACCCCAAGTAGGGGAAAGAGACTCA     | AGCTTTACAGAGAGAA          |                 |        | 1839 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 503  | LeuProAspGlyTyrlGln    | ThrArgValaGlyGln              | ArgGlyLeuLysLeuSerGly     | ValGln          |        | 522  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1840 | AAGGAAGAGAGCAAT        | TGGAAGAAGCATTTGGAAGAGACCC     | CCCAAGTACTACTAT           |                 |        | 1899 |
|   |   |      |                        |                               |                           |                 |        |      |
| D | b | 523  | LysGlnArgValAlaIle     | IaIaIaArgThrIleLeuLysLys      | ProSerIleLeuValPheAsp     |                 |        | 542  |
|   |   |      |                        |                               |                           |                 |        |      |
| Q | y | 1900 | GAAGCTACTTCATGCT       | TAGATTCGATTACTGAAGACATTTCTTG  | TGCTGACATGAAGAT           |                 |        | 1959 |
|   |   |      |                        |                               |                           |                 |        |      |

| ID                                 | Q9ZDM0  | PRELIMINARY;  | PR7;          | 609 AA. |
|------------------------------------|---|---|---------------|---------|
| AC                                 | Q9ZDM0  |   |               |         |
| DT                                 | 01-MAY-1999   | (TREMBLrel. 10, Created)  |               |         |
| DT                                 | 01-MAY-1999   | (TREMBLrel. 10, Last sequence update)                           |               |         |
| DT                                 | 01-JUN-2002   | (TREMBLrel. 21, Last annotation update)                         |               |         |
| DE                                 | Mitochondrial transporter ATW1 precursor (ATW1).                    |   |               |         |
| GN                                 | RP205.  |   |               |         |
| OS                                 | Rickettsia prowazekii.  |   |               |         |
| OC                                 | Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;         |   |               |         |
| OC                                 | Rickettsiaceae; Rickettsiellae; Rickettsia.                         |   |               |         |
| OX                                 | NCBI_TextID=782;  |   |               |         |
| RN                                 | [1]   |   |               |         |
| RP                                 | SEQUENCE FROM N.A.  |   |               |         |
| RC                                 | STRAIN=MADRID E;  |   |               |         |
| RC                                 | MEDLINE=99039499; PubMed=9823893;                                   |   |               |         |
| RA                                 | Andersson S.G.E., Zomrodipour A., Andersson J.O.,                   |   |               |         |
| RA                                 | Siderheltz-Porten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K., |   |               |         |
| RA                                 | Eriksson A.-S., Winkler H.H., Kurland C.G.; and the origin of       |   |               |         |
| RT                                 | "the genome sequence of Rickettsia prowazekii and the origin of     |   |               |         |
| RL                                 | mitochondria."  |   |               |         |
| RL                                 | Nature 396:133-140(1998).   |   |               |         |
| CC                                 | -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY |   |               |         |
| CC                                 | (ABC TRANSPORTERS).   |   |               |         |
| DR                                 | EMBL; AJ235270; CAA14670.1; -                                       |   |               |         |
| DR                                 | HSSP; P13569; INBD.   |   |               |         |
| DR                                 | InterPro; IPR003593; AAA ATPase.                                    |   |               |         |
| DR                                 | InterPro; IPR001140; ABCtransportM.                                 |   |               |         |
| DR                                 | InterPro; IPR003439; ABC transportr.                                |   |               |         |
| DR                                 | Pfam; PF00664; ABC_membrane; 1.                                     |   |               |         |
| DR                                 | Pfam; PF00005; ABC_tran; 1.   |   |               |         |
| DR                                 | ProDom; PD00006; ABC_transport; 1.                                  |   |               |         |
| DR                                 | SMART; SM00382; AAA; 1.   |   |               |         |
| DR                                 | PROSITE; PS00211; ABC_TRANSPORTER; 1.                               |   |               |         |
| KW                                 | ATP-binding; Transport; Complete proteome.                          |   |               |         |
| SQ                                 | SEQUENCE 609 AA; 68668 MW; 1126067A0F0D8243 CRC64;                  |   |               |         |
| Alignment Scores:                  |   |   |               |         |
|                                    | Pred. NO.:  | 4,21e-92  | Length:       | 609     |
|                                    | Score:  | 1446.50   | Matches:      | 294     |
|                                    | Percent Similarity:   | 68.67%  | Conservative: | 118     |
|                                    | Best Local Similarity:  | 49.00%  | Mismatches:   | 170     |
|                                    | Query Match:  | 34.00%  | Indels:       | 18      |
|                                    | DB:   | 16  | Gaps:         | 6       |
| AF133659 (1-2345) x Q9ZDM0 (1-609) |   |   |               |         |
| QY                                 | 370   | ATGCTTCTTATGTGTGGCCCAAGACGAGCCAGATCTAGAGAGTTCAGTTCG             | 4228          |         |
| DB                                 | 23  | LeuLeuThrTyrLeuIleuTrpSerIysAspPhe--AspIleArgLeuArgIleValThrSer | 41            |         |
| QY                                 | 430   | CTGGGATTTTGGGGTGGGCAAGGCAATGATATTTGGTGTCCCTCATGTTAAATAT         | 489           |         |
| DB                                 | 42  | LeuIleCysLeuValIleAlaIysValIleLeuIlePheValProlIleValTyrLysTyr   | 61            |         |
| QY                                 | 490   | GCTGTGACAGCCTTCAACCAAGATGTGGGAAACATGCTGAACCTGAAGTGCACCAAT       | 5498          |         |
| DB                                 | 62  | IleIleAspGlyLeuAsnGln-----AsnLeuSer-----                        | 71            |         |
| QY                                 | 550   | ACAAGTTCACACATGGCAACGACGACTTCTGATTTGGCTATGGTGTATCAAGAGCTGAGCT   | 6099          |         |
| DB                                 | 72  | -----LeuSerValLeuIleGlyValIleIleIleGlyTyrGlyGlyThrIysIleLeuAla  | 89            |         |
| QY                                 | 610   | GCTTTTTTTAAAGAGTTCCAAATGACAGTATTTGGCAAGGTAGCCCAAGATTCATCCGA     | 6658          |         |
| DB                                 | 90  | GlnIlePheSerGlnLeuArgGlnIleIlePheSerIysValGlyCysGlnAlaThrArg    | 1099          |         |
| QY                                 | 670   | AGAATAGCCAAATATGCTTTCTTCATCTTCAACACACCTGATCTGGGTTTTCACCTGAGC    | 7299          |         |
| DB                                 | 110   | LeuValAlaIleAsnValPheIysHisMetHisAsnLeuSerMetArgPheHisIleThr    | 1299          |         |
| QY                                 | 730   | AGACAGACGGAGGCTTATATCATAGGCTATTATGACAGAGAAACAAGGGGTATCAGTTTGTG  | 7899          |         |
| DB                                 | 130   | ArgValThrGlyIysIleSerArgSerIleGlnArgGlyThrIysGlyIleGlnAlaVal    | 1499          |         |





Db 33 LeuArgMetArgValValTrrPalatThrPheThrLeuValLeuSerIleValLeuIle 52  
 QY 466 GTGGTTCCTTCATGTTAAATATGCTGTAGACAGCTCAACAGATGCGGAAACATG 525  
 Db 53 LeuValProTyrPhePheIleTrrValThrAsnAlaLeuAsn-----GlyGlnLeu 69  
 QY 526 CTGAACCTGAGTGAACCAATATGACCTTGCACACAG-----GCACAGCACTT 576  
 Db 70 -----HisAlaProTyrTrrIleProValIleLeuValGlyAlaValMetLeu 85  
 QY 577 CTGATGGCTATGGTGTATCAAGAGCTGAGAGCTCTTTTATCAAGAACTCGAATATGCA 636  
 Db 86 ValLeuAlaTyrAsnAlaIleValIleValGlnIleValGlnLeuAsnGlnLeuArgAspAla 105  
 QY 637 GTATTGGCAAGGTAGCCCAAGATTCATCCGAAGAAATGCCAAATAATGCTTCTTCAT 696  
 Db 106 LeuPheAlaSerValGlyGlnTrrValAlaValArgIleLeuAlaTrrIleValHis 125  
 QY 697 CTTCACAACTCGATCTGGGTTTCACTGAGCAGACAGAGCGGAGCTTATCTTAAGCT 756  
 Db 126 MetHisGlnLeuSerLeuValArgPheHisLeuGlnIleArgTrpGlyLeuSerArgVal 145  
 QY 757 ATTGACAGAGCAAGCAAGGGGTATCAGTTTGTCCGAGTGTCTTGTATTAATCTTCTT 816  
 Db 146 IleGlnArgGlyThrIleGlyIleGlnThrIleValArgPheThrIleLeuAsnThrLeu 165  
 QY 817 CCATCATGTTTGAAGTATGATGCTGTGAGTGTGTTGTATTAATAATGCGTCCAG 876  
 Db 166 ProThrIleLeuGlnIlePheAlaLeuThrAlaValIlePheAlaPheAlaTrrGlyLeuSer 185  
 QY 877 TTTCCTTGTGTAACCTTGGCAACCTTGATACATACAGCATTCACAGTTGCACTGCA 936  
 Db 186 TyrLeuValValAlaAlaIleThrValTrrLeuTrrIleThrPheThrIleArgAlaSer 205  
 QY 937 CGGAGGAACTAGATTAGAAATAGAAATAGCAAGCAAGTATGACAGTAATGCT 996  
 Db 206 AspTrrPheGlnIleAsnIleIleArgGlnMetCysAspSerPheThrAspAlaAsnThrLys 225  
 QY 997 GCTATAGACTCACTGCTGAATATTAAGAACTGTGAAGTATTTAATAAGAAAGTATGAA 1056  
 Db 226 AlaIleAspSerLeuLeuAsnPheGlnThrValIleTrrPheGlyAsnGlnAlaMetGln 245  
 QY 1057 GCACAGAGATATGATGATTTTGAAGACGTATGAGACTGCTTATTCGAAAGTACCTCT 1116  
 Db 246 AlaTyrArgPheAspGlyAlaMetAlaArgTrrGlyLysAlaAlaIleThrGlnThrTrrThr 265  
 QY 1117 ACTGTGGCTATGCTGAACCTTGTGCAAGGCTATTTTTCAGTGGCTTAAACGCTATA 1176  
 Db 266 SerLeuGlnTrrPheAsnPheGlnIleValAlaIlePheGlyAlaGlyMetAlaIleVal 285  
 QY 1177 ATGTGTCTGCGCACTGAGGAAATGTGAGAGTACCTTACTGTGGAGATCTAGTAATG 1236  
 Db 286 MetValMetSerAlaMetGlnValGlnIleGlyThrGlnSerLeuGlyAspPheValPhe 305  
 QY 1237 GTGAATGAGCTGCTTTTCAAGCTTTCATTAACCTTGAACCTTTCGGAACGTATATAGA 1296  
 Db 306 IleAsnAlaLeuLeuMetGlnLeuSerIleProLeuAsnPheIleGlyPheIleTrrArg 325  
 QY 1297 GAGACTGAGACGCTCATAGATATGAACACCTGTTTACTTACTCAAGATGAGAACACC 1356  
 Db 326 GlnIleArgGlnIleGlnLeuThrAspIleGlnGlnMetPheLeuAspValIleGln 345  
 QY 1357 CAATTTAAAGACAAAGTATGATGCTCCCTTCAGATCAACACAGACAGTACCGTG 1416  
 Db 346 GlnValIleAspIleProGlyAlaProAlaLeuLysIle-----AspSerGlyAlaIle 363  
 QY 1417 GCCTTTGATATGTCATTTTGAATCATTTGAGGGCCAGAAAGTCTTATGTAATATCC 1476  
 Db 364 SerPheLysAspValHisPheAlaTrrAspProGlnIleArgProIleLeuArgIleSer 383  
 QY 1477 TTGGAAGTCCCTGAGGAAAGAAAGTGGCATTTAGAGAGTATGGTGTACAGGAAAGC 1536  
 Db 384 PheAspValProAlaGlyLysThrValAlaIleValGlyProSerGlyAlaGlyLysSer 403

QY 1537 ACAATAGTAGGCTATTATTTCGCTTCTATAGCCTCAAAAAGGTACCATTTATCTTGCT 1596  
 Db 404 ThrIleSerArgLeuLeuPheArgPheTrrAspIleGlnSerGlySerIleThrIleAsp 423  
 QY 1597 GGTCAAAATATATACAGATGTGAGCCTTGGAAGCCTTGGAGGGCAGAGGTGACTACT 1656  
 Db 424 GlyGlnAspValArgAspValIleGlnIleSerLeuArgLysPheIleGlyMetValPro 443  
 QY 1657 CAGGATGCTGCTCTTCCTCAATATCTATTATTAACAACCTTATATGAAACATCAGT 1716  
 Db 444 GlnAspThrValLeuPheAsnAspThrIleAlaTrrAsnIleArgTrrGlyArgThrAsp 463  
 QY 1717 GCTTTCACCTGAGAGATGATGACGTGCGCAAAATTTAGCTGCACTCATGATGCAATCTT 1776  
 Db 464 AlaSerGlnIleAspValGlyLysAlaIleGlnLeuAlaGlnIleAlaIleGlyPheIleLys 483  
 QY 1777 CGAATGCCATGATGATGACACCCAGTGGGGAAGAGACGCTCAAGCTTTCAGAGGA 1836  
 Db 484 HisLeuProAspGlyTrrLysSerMetValGlyGlnArgGlyLeuLysSerGlyGly 503  
 QY 1837 GAAAGCAAGAGTAGCAATTTGCAAGAGCCATTTTGAAGACCCCGATCATCTCTAT 1896  
 Db 504 GlnLysGlnArgValAlaIleAlaIleArgThrIleLeuLysAlaProProIleLeuIleLeu 523  
 QY 1897 GATGAAGCTACTTCATGCTTATGATTCGATTACTGAAGACATATTCTTGTCCTCAAG 1956  
 Db 524 AspGlnIleThrSerAlaLeuAspThrAlaThrGlnIleGlnIleGlnSerAlaLeuAsp 543  
 QY 1957 GATGTGCTCAACACAGAACTTCTATTTCATTTGACACAGATTTGCAACAGTGTGAT 2016  
 Db 544 IleValSerArgGlyTrrThrIleValIleAlaHisArgLeuSerThrValIleGly 563  
 QY 2017 GCAGAGAAATCATGCTGCTGATGATGAGGTAGGTAGCCAGGTGTACCCACATGCT 2076  
 Db 564 AlaAspGlnIleIleValLeuLysAspGlyLeuIleAlaGlnArgGlyThrHisArgHis 583  
 QY 2077 TTGCTTGTCAACCTCATATGATATTCATTCAGAAATGTGGCATCACAGACGCTGTG 2136  
 Db 584 LeuLeu---AspGlnLysGlyLeuTrrValaSerMetTrrAspArgGln----- 598  
 QY 2137 CAAGACCATATATACCCCAATAGTGGAAGCAAGAAAGAAATATTCACAAAGAGAGGA 2196  
 Db 599 -----ArgIleAlaSerGlnAlaGlnArg 607  
 QY 2197 AGAAGAACTACAGAA 2214  
 Db 608 LeuArgIleValArgGln 613  
 RESULT 12,  
 08Y258  
 AC 08Y258; PRELIMINARY; PRT: 592 AA.  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
 DE Probable composite ATP-binding transmembrane ABC transporter protein.  
 GN RSC0478 OR RSC04417.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_Taxid=305;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM1100;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Arlat M., Billault A., Brotier P., Camus J.C., Catcollco L.,  
 RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.;



RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 DR EMBL: AL646059; CAD14006.1; --  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001140; ABCtransporter.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR Pfam: PF00664; ABC\_membrane.1.  
 DR Pfam: PF00005; ABC\_tran.1.  
 DR ProDom: PD00006; ABC\_transportr.1.  
 DR SMART: SM00382; AAA\_I.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER.1.  
 DR Complete proteome.  
 SEQ SEQUENCE 592 AA; 66708 MW; 3EFAB0EA350B9B0 CRC64;

## Alignment Scores:

| Pred. No.:             | 8,22e-69 | Length:       | 592 |
|------------------------|----------|---------------|-----|
| Score:                 | 1399.00  | Matches:      | 277 |
| Percent Similarity:    | 67.11%   | Conservative: | 123 |
| Best Local Similarity: | 46.48%   | Mismatches:   | 178 |
| Query Match:           | 32.89%   | Indels:       | 18  |
| DB:                    | 16       | Gaps:         | 6   |

AF133659 (1-2345) x Q8Y258 (1-592)

```

QY 361 ATAAAGCATGCTTTCTATGTCGCCCAAGACAGGCCAGATCTACGACTGAGTT 420
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 8 IleArgSerLeuLeuProTyrLeuThrPala-----TyrLysThrPrgVal 22

QY 421 GCCATTTGGCTGGATTTTGGGTGGTGAAGCCAGTAATATTGGTCCCTTCATG 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 23 AlaLeuAlaLeuSerPheLeuIleAlaIleValAlaIleAsnLeuGlyValPrometVal 42

QY 481 TTTAATATGCTGTGACAGCCTCAACACAGATGCGGAAACATGCTGACCTGAGTAT 540
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 43 MetLysArgLeuIleAspAlaMetAsn-----ValSerProThrAsp 56

QY 541 GCACCAATATACAGTTGCAACCATGGCAACAGAGCTTCGATGGCTATGGTATCAACA 600
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 57 ---ProValGlaLeuLeuValAlaProValGlyIleIleLeuGlyTyrGlyLeuLeuArg 75

QY 601 GCTGAGCTGCTTTTATTAACGAAGTTCGAATGCAATTTTGGACAGTACCCAGAAAT 660
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 76 LeuSerThrSerLeuPheSerGluLeuArgGluIleLeuPheSerLysValThrGluSer 95

QY 661 TCAATCCGAGAAATGCCCAAAATGCTTTCATCTTCATCTTCACAACTGATCGGTTT 720
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 96 SerValArgThrLeuAlaLeuGluIleValPheArgHisLeuHisAlaLeuSerLeuArgPhe 115

QY 721 CACCTGACAGACAGACGGAGCTTATCTAAGCTATGACAGACAGAAACAGGAGTATC 780
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 116 HisLeuGluIleArgGlnThrGlyMetSerArgAspIleGluIleGlyThrArgGlyIle 135

QY 781 AGTTTTCCTGAGTCTTGGTATTAATCTTCCATCATGATGTTGAAGTATGCTT 840
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 136 GlnSerLeuIleSerLysSerLeuTyrSerIleLeuProThrLeuValGluValGlyLeu 155

QY 841 GTCAGTGGCTTTTGTATATCAAAATGCGGTCGCCAGTTGCTTGGTACCTTGGACA 900
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156 ValIleThrTyrPhePheValIleLysTyrAspAlaThrPheAlaLeuIleThrPheCysAla 175

QY 901 CTGTGATACATACAGACATTCACAGTTGACATGACAGCGGTGAGAACTAGATTGAATA 960
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 176 LeuValSerTyrIleValPheThrValThrValThrAsnThrArgThrHisPheArgArg 195

QY 961 GAAATGAACAAACAGATATATGATGAGTAATGCTGCTATAGACTCACTGCTGAATAT 1020
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 196 ArgMetAsnGluLeuAspSerArgIleAsnGlnLysAlaIleAspSerLeuLeuAsnPhe 215

QY 1021 GAAATGGAAGTATTTTAAATAGAAAGATATGAAGACAGACAGATATGATTTTGG 1080
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 216 GlnThrValIleLysTyrPheGlyAsnGluGluTyrGluThrArgArgGlyArgAspIleAsnLeu 235

QY 1081 AAGACGTATGACAGTCTTCAATGAAAGTAACTCTACTCTGCTATGCTGAATTTGGT 1140

```

```

Db 236 ArgLysTyrArgAlaAlaAlaIleArgSerGlnHisSerLeuSerLeuLeuAsnPheGly 255
   ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1141 CAAGTGCATATTTTCAGTGCCTTTTAAACAGCTTAATGCGTCCGACAGCAGGAAT 1200
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 256 GlnGlnLeuIleValAlaValAlaLeuIleLeuIleLeuTyrArgAlaThrGlnGlyVal 275

QY 1201 GTGCGAGTACCTTACTGTGTGAGATCTAGTAATGAGTGAATGAGCTGTTTTCAGCTT 1260
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 276 AlaAlaGlyHisMetThrIleGlyAspLeuValLeuValAsnThrLeuMetLeuGlnIle 295

QY 1261 TCATTTACCTTGAACCTTTCTGGGAACTGTATATGAGAGCTAGACAAACACTATAGAT 1320
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 296 TyrIleProLeuAsnPheLeuGlyValIleTyrArgGluLeuLysGlnAlaValThrAsp 315

QY 1321 ATGAACACCTTGTACTCTGATCAAGTGAACGCCCAATTAAGCAAAAGTATGAGCA 1380
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 316 MetAspArgMetPheLysLeuLeuHisThrAsnArgGluValAlaAspArgProAspPhe 335

QY 1381 TCTCCCTTCAGATCACACACAGACAGTACCTGCGCTTGTGAATGTGCAATTTTGA 1440
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 336 GlnProLeuAlaVal-----ArgAlaGlyGluValArgPheAlaHisValAspPheGly 353

QY 1441 TACATTGAGGGCCAGAAAGTCTTATGGAATATCTTGAATCTTGAAGTCCCTGCAGAAAGAA 1500
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 354 TyrGluSerAsnArgGlnIleLeuPheAspValAspPheThrIleProAlaGlyThrThr 373

QY 1501 GTGCGCATTTAGAGAGGTAGTGGTCAAGGAAAGCAATAGAGAGCTATTTTTCGC 1560
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 374 ThrAlaValAlaGlyGlnSerGlySerGlyLysSerThrLeuAlaArgLeuLeuPheArg 393

QY 1561 TTCTATGACCTCAAAAGGTAGACATTTATCTGCTGTCGTCGTCGTCGTCGTCGTCGTCG 1620
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 394 PheTyrAspAlaThrSerGlyAlaIleGlnIleAspGlyGlnAspValArgAspValThr 413

QY 1621 CTGAAAGCTTCGAGAGGCGAGTGGAGTGTACTCAGATGCTGCTCTTTCATATAT 1680
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 414 GlnAlaSerValArgAlaAlaIleGlyIleValProGlnAspThrValLeuPheAsnAsp 433

QY 1681 ACTATTTTATCAACCTTTATATGGAACATCATGCTGATCACTGCGGAGAGTATGATCA 1740
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 434 SerIleTyrIleAsnIleAlaTyrGlyArgProAspAlaThrArgGluGluValIleAla 453

QY 1741 GTGCGCAATATAGCTGCTCATGATGCAATTTCTGCAATGCGACATGATATGACACC 1800
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 454 AlaAlaArgAlaAlaGlnIleHisGlyPheValGluSerLeuProGluGlyTyrAspPhe 473

QY 1801 CAAGTAGGGAACAGACCTCAACCTTTTCAGAGAGAGAAAGCAAGATGCAATTTGCA 1860
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 474 ProValGlyGluArgGlyLeuLysLeuSerGlyGlyGluLysGlnArgValAlaIleAla 493

QY 1861 AGACCATTTTGAAGAGACCCCCAGTCACTACTCATATGAAAGTACTCTCACTGAT 1920
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 494 ArgThrLeuLeuLysArgProProIleLeuValPheAspGlnAlaThrSerAlaLeuAsp 513

QY 1921 TCGATTTACTGAAGACATTTCTGTGTCGCATGAAGATGCTCAACACAGAACTTCT 1980
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 514 SerArgThrGlnHisAlaIleGlnGluLeuMetArgLeuAlaGlnAsnHisThrThr 533

QY 1981 ATTTTCATGACACACAGATTTTCAACAGTGGTATGATGACATGAATATGCTTGGAT 2040
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 534 LeuValIleAlaHisArgLeuSerThrIleValGlyAlaHisGlnIleLeuValMetGlu 553

QY 2041 CAGGTATAGTACCGCAAGCTGTATCCACAGATGTTGCTTGTGTAACCCCTATAGTATC 2100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 554 HisGlyArgIleIleGluArgGlyThrHisAlaSerLeuLeu---ArgAlaGluGlyArg 572

QY 2101 TATTCAGAAATGTGCGATACACAGAGC-----AGCCGTGTGACG 2139
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 573 TyrIleGlnMetThrPargMetGlnAlaArgGluProGluArgValGln 588

RESULT 13
Q984W3

```





QY 1339 CTACTGAGGTAGACACCAATTAAAGACAAGATGATGCGATCCCTTCAGATCACA 1338  
 ||||| : : : : : ||||| : : : : :  
 Db 342 LeuLeuGluValIGluValAGluValThrAspArgProAspAlaValSPLeuAlaValAGly 361  
 QY 1339 CCACAGACAGCTACCGTGGCCCTTGATATATGTCATTTTGAATTCATTGAGGCCAGAAA 1458  
 ||||| : : : : : ||||| : : : : :  
 Db 362 Pro-----GlyAlaIleSerPheArgValHisPheAlaTyAspProGluArgPro 379  
 QY 1459 GTCCCTAGTGAATATCTTTGAAGTCCCTGAGAAAAGAGGCCATTGTAGAGCT 1518  
 ||||| : : : : : ||||| : : : : :  
 Db 380 IleLeuysGlyValSerPheAspValProAlaGlyStrValAlaIleValGlyPro 399  
 QY 1519 AGTGGGTGAGGAAAAGACAAATAGTGAAGCTATTTATTCCTTCTAGAGCTCAAAAG 1578  
 ||||| : : : : : ||||| : : : : :  
 Db 400 SerGlyAlaGlySerThrIleSerArgLeuLeuTyArgPheTyAspIleGlnGlu 419  
 QY 1579 GGTAGCATTTATCTTGGCTGGTCAAAATATACAGATGTGAGCCCTGGAAAGCTTGGAGG 1638  
 ||||| : : : : : ||||| : : : : :  
 Db 420 GlyAlaValThrIleAspGlyGlnAspIleArgAspValThrGlnIleSerLeuArgSer 439  
 QY 1639 GCAGTGGAGAGTACCTCAGATGCTGCTCTCCATATACATTTATTTACAACTC 1698  
 ||||| : : : : : ||||| : : : : :  
 Db 440 MetIleGlyMetValProGlnAspThrValLeuPheAspAspThrLeuAlaTyAsnIle 459  
 QY 1699 TTATATGAGAAACATCAGTGTCTTCACTGAGAGAGTGTAGTGCAGTGCAGAAATTAAGCTGA 1758  
 ||||| : : : : : ||||| : : : : :  
 Db 460 ArgTyArgIleArgProSerAlaThrAspGluLeuLeuysAlaAlaAspAlaValGln 479  
 QY 1759 CTTGATGATGCAATCTTTCGAATGCGACATGATGTGACACCCCAATGGGAAAGAGGA 1818  
 ||||| : : : : : ||||| : : : : :  
 Db 480 IleSerAlaPheIleGlyLeuProAspGlyTyAlaThrMetValGlyGluArgGly 499  
 QY 1819 CTGAGCTTTAGAGAGGAGAAAGAAAGATGAGATTTGAGAGCCATTTTGAAGAC 1878  
 ||||| : : : : : ||||| : : : : :  
 Db 500 LeuysLeuSerGlyGlyGlnGlnArgValAlaIleAlaArgThrIleLeuysAla 519  
 QY 1879 CCCCAGTCACTCTATGATGAGAGCTCACTCATGCTAGATTCGATTAAGAGACT 1938  
 ||||| : : : : : ||||| : : : : :  
 Db 520 ProIleLeuIleLeuAspGluAlaThrSerAlaLeuAspThrIleThrGlnGln 539  
 QY 1939 ATCTTGGTGCATGAGATGTGTCACAAAGCAAACTTATTTTCATTTGACACAGA 1998  
 ||||| : : : : : ||||| : : : : :  
 Db 540 IleGlnSerAlaLeuAspIleValSerTyAsnArgThrThrLeuValIleAlaHisArg 559  
 QY 1999 TTGTCACAGTGTGATGAGATGAATCATGCTTGTGATTCAGAGGTAAGTACCGCA 2058  
 ||||| : : : : : ||||| : : : : :  
 Db 560 LeuSerThrValIleHisAlaAspGluIleValLeuysGlnGlyLeuIleAlaGln 579  
 QY 2059 CGTGTACCCACATGCTTGTCTTCAACCTCATAGTATCTATTCAGAAATGTGCGAT 2118  
 ||||| : : : : : ||||| : : : : :  
 Db 580 ArgGlyThrHisAlaSerLeuMetAlaHis---AspGlyLeuTyAlaSerMetTyrSer 598  
 QY 2119 ACACAGACAGCCGTGTGACAGAACCATGATTAACCCCAATGGAGCAAGAAAGAAAT 2178  
 ||||| : : : : : ||||| : : : : :  
 Db 599 ArgGln-----ArgGluAla 603  
 QY 2179 ATATCCAAAGAGAGAAAGAAAGAAACTACAGAA 2214  
 ||||| : : : : : ||||| : : : : :  
 Db 604 IleArgAlaGlnGluMetLeuArgHisValArgGln 615  
 RESULT 15  
 Q92R12  
 ID Q92R12 PRELIMINARY; PRT; 627 AA.  
 AC Q92R12;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DE Probable ABC transporter ATP-binding transmembrane protein.  
 GN R01120 OR SMC00550.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC Rhizobiaceae; Sinorhizobium.

OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
 RA Pohl T., Portecelle D., Puehler A., Purnelle B., Ramsperger U.,  
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 Sinorhizobium meliloti strain 1021.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 DR EMBL; AL591786; CAC45699.1; -.  
 DR InterPro; IPR001140; ABCcramptrTM.  
 DR InterPro; IPR003439; ABC transporter.  
 DR Pfam; PF00654; ABC\_membrane; 1.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 627 AA; 70055 MW; 3C0BEA68107E451E CRC64;  
 Alignment Scores:  
 Pred. No.: 9.95e-87 Length: 627  
 Score: 1369.00 Matches: 280  
 Percent Similarity: 65.04% Conservative: 120  
 Best Local Similarity: 45.53% Mismatches: 187  
 Query Match: 32.18% Indels: 28  
 DB: Gaps: 5  
 AF133659 (1-2345) x Q92R12 (1-627)  
 QY 379 TATGTGTGGCCCAAGACAGCCAGATCTACAGACTAGTGGCATTTGCCGTGATTT 438  
 ||||| : : : : : ||||| : : : : :  
 Db 24 TyMetTrpProAlaAspArgIleAspLeuIleuArgValAlaThrAlaThrValIle 43  
 QY 439 TTGGGTGTGCAAAAGCGCATGATATGTCCTTCCCTTCAATGTTTAAATATGCTGTGAG 498  
 ||||| : : : : : ||||| : : : : :  
 Db 44 LeuValIleAlaIleAlaValAlaValIleValIleValProTyRhePheIleThrAlaThrAsp 63  
 QY 499 AGCCTCAACAGATGTGCGGAAACATGCTGAACCTGAGTATGACCAATACAGTT--- 555  
 ||||| : : : : : ||||| : : : : :  
 Db 64 AlaleuAsn-----AsnIleProAspAlaLeuGlyPheLeuPro 76  
 QY 556 -----GAAACCATGCGAAGACAGATCTGATTTGCGTATGATGATGATGATGATG 605  
 ||||| : : : : : ||||| : : : : :  
 Db 77 GlnPheLeuThrGlyAlaValMetLeuValIleuAlaTyAsnIleuAlaArgLeuGln 96  
 QY 610 GCTTTTATTAAGAGATGCAAAATGACATTTTGGCAAGGAGGACCGAATTAATCGGA 669  
 ||||| : : : : : ||||| : : : : :  
 Db 97 AlGlyLeuAsnGlnIleuAlaGlyAspAlaLeuPheAlaSerValGlyGlnHisAlaValArg 116  
 QY 670 AGAATAGCCAAATAATGTCTTCTCATCTTCACAACTGATGCTGGTTCACSTGAGC 729  
 ||||| : : : : : ||||| : : : : :  
 Db 117 GlnLeuAlaTyIleTyStrPheValHisMetHisGlnLeuSerLeuArgPheHisLeuGln 136  
 QY 730 AGACAGACGGGAGCTTTATGTTAAGCTATGACAGAGAAACAAGGGTATACATTTGTC 789  
 ||||| : : : : : ||||| : : : : :  
 Db 137 ArgArgThrGlyIleuSerArgIleIleGluArgGlyThrIleArgGlyIleGluThrIle 156  
 QY 790 CTGAGTCTTGGATATTATCTTCTCCATCTTCACAACTGATGCTGGTTCACSTGAGC 849  
 ||||| : : : : : ||||| : : : : :  
 Db 157 ValArgPheThrIleLeuAsnSerValProThrLeuIleGluPheLeuThrAlaVal 176  
 QY 850 GTTTTGTATTACAAATGCGGTGCGAGTTTGTCTTGTGTAACCTTGGAAACATTTGATACA 909  
 ||||| : : : : : ||||| : : : : :  
 Db 177 IlePheTrpTrpGlyTyArgIlePheThrTyLeuLeuValThrAlaValThrValTrpLeu 196  
 QY 910 TACACAGCATTCACAGTTGACATGACAGCGGTGAGAACTGATTTAGAAATGAATGAC 969  
 ||||| : : : : : ||||| : : : : :  
 Db 197 TyIleTrpPheThrValArgAlaSerAspTrpArgIleAlaIleArgArgSerMetAsn 216

QY 970 AAAGCAGATTAATGATGACGTAATCTGCTATAGACTGCTGTAATATGAACTG 1029  
Db 217 AsperAspThrAspAlaAsnThrLysAlaIleAspSerLeuLeuAsnPheGluThrVal 236  
QY 1030 AAGTATTTTAAATGAAGAATATAGACACAGAGATATGATGATTTTGAAGCGTAT 1089  
Db 237 LysTyrPheGlyAsnGluMetGluAlaLysArgPheAspLysSerMetGluArgTyr 256  
QY 1090 GAGACTGCTTCACTGAAAGTACCTCTGCTGCTATGCTGACCTTTGGTCAAGTGT 1149  
Db 257 GluArgAlaAlaThrGlnAlaIlePheSerLeuGlyTyrPheAsnPheGlyGlnAlaLeu 276  
QY 1150 ATTTTCAGTGTGCTTTTAAACGCTTAATGTGCTGCGCAGTCAAGAAATGTCAGGT 1209  
Db 277 IlePheGlyAlaGlyThrAlaValMetMetThrIleSerAlaLeuAlaValGlnArgGly 296  
QY 1210 ACCCTTACTGTGAGATCTAGTAATGTGAATGAGCTGCTTTTTCAGCTTCAATACC 1269  
Db 297 GluGlnThrIleGlyAspPheValPheIleAsnAlaMetLeuIleGlnLeuAlaIlePro 316  
QY 1270 CTGAACCTTCTGCGAAGCTGATATATAGAGACTGACAGCAAGCACTGATATGAACAC 1329  
Db 317 LeuAsnPheIleGlyPheValTyrArgGluIleArgGlnGlyLeuThrAspIleGluHis 336  
QY 1330 TTGTTTACTTACTCAAGGTAGACACCCAAATTAAGACAAAGTATGATGATCCCTT 1389  
Db 337 MetPheAspLeuAspValArgAlaGluValAlaAspArgProAsnAlaLysGluLeu 356  
QY 1390 CAGATCACACACAGACAGCTACCGTGGCTTGTATATGTCATTTGATACATTGAG 1449  
Db 357 LysIle-----GluArgGlyAlaIleAlaPheLysAsnValGlnPheAlaTyrAspPro 374  
QY 1450 GGCAGAAAGTCTAGTGAATATCCTTGAAGTCCCTGACAGAAAGTGGCCATT 1509  
Db 375 AlaArgProIleLeuLysGlyIleSerPheGlnValProAlaGlyLysThrValAlaVal 394  
QY 1510 GTAGAGGTAGTGGTCAAGGAAAGCAACAATAGTGAAGCTTATTTGCTTCTATGAG 1569  
Db 395 ValGlyProSerGlyAlaGlyLysSerThrLeuSerArgLeuLeuTyrArgPheTyrAsp 414  
QY 1570 CCTCAAAAGGTAGCATTTATCTGCTGTCAAAATATACAAGATGTGACCTGAAAAC 1629  
Db 415 ValGlnGlnGlySerIleThrIleAspGlyGlnAspValArgAspValThrGlnLysSer 434  
QY 1630 CTTGGAGGAGGAGTGGAGTGTACTGATGATGCTGCTCTTCCATAATACTATTAT 1689  
Db 435 LeuArgSerValIleGlyMetValProGlnAspThrValLeuPheAsnAspThrIleAla 454  
QY 1690 TACAACCTCTTATGAAACATCATGCTTCACTGAGAGAGTATGACAGTGGCAAAA 1749  
Db 455 TyrAsnIleArgTyrGlyArgValThrAlaSerAlaGluValGluAlaAlaGlu 474  
QY 1750 TTAGCTGACCTTCATGATGCAATTTCTTCGATGCAATGATGATGACACCAAGTGGG 1809  
Db 475 AlaIleGlnIleAlaAspPheIleArgAspLeuProGluGlyPheArgAlaMetValGly 494  
QY 1810 GAACGAGGACTCAAGCTTTCAGAGAGAAAGCAAGAGTAGCAATTGCAAGCCATT 1869  
Db 495 GluArgGlyLeuLysLeuSerGlyGlyGlnLysGlnArgValAlaIleAlaArgThrIle 514  
QY 1870 TTGAAGACCCCTCAGTCTACTATGATGAGTGAAGTACTTCACTGTTAGATTGATTA 1929  
Db 515 LeuLysAlaProPheIleLeuIleLeuAspGluAlaThrSerAlaLeuAspThrLysThr 534  
QY 1930 GAAGAGACTATTTCTGTGTCATGAGAGATGTGTCAAACACAGAACTTATTTTCATT 1989  
Db 535 GluGlnGluIleGlnAlaAlaLeuAspIleValSerArgAsnArgThrLeuValIle 554  
QY 1990 GCACACAAATTTGTCAACAGTGTGATGACAGATGAATGAAATGATGTCAGGTAAG 2049  
Db 555 AlaHisArgLeuSerThrValIleAsnAlaAspGluIleIleValLeuLysAspGlyVal 574  
QY 2050 GTAGCCGAACGTGTACCAACATGATGTTGCTTGAACCTCATAGTATCATTCAGAA 2109

Db 575 IleAlaGluArgGlyThr---HisGlyGluLeuIleAspArgGlyGlyLeuTyrAlaSer 593  
QY 2110 ATGTGGCATATACACAGAGAGCCGTGTGACAGAACATGATTAACCCCAATGGAGCAAG 2169  
Db 594 MetTyrSerArgGln----- 598  
QY 2170 AAGAAATATATCCAAAGAGAGAAAGAAAGAACTACAAAGA 2214  
Db 599 ArgGluAlaThrGlnAlaGluGluGlnLeuLysArgValArgGlu 613

Search completed: February 15, 2003, 05:58:06  
Job time : 114 secs